

Spatially regularized compressed sensing of diffusion MRI data

Oleg Michailovich, Yogesh Rathi and Sudipto Dolui *

September 21, 2010

Abstract

Despite the relative recency of its inception, the theory of compressive sampling (aka compressed sensing) (CS) has already revolutionized multiple areas of applied sciences, a particularly important instance of which is medical imaging. Specifically, the theory appears to provide an answer to the important problem of optimal sampling in MRI, with an ever-increasing body of works reporting stable and accurate reconstruction of MRI scans from the number of spectral measurements which would have been deemed unacceptably small as recently as five years ago. Reducing the number of MR measurements per scan comes to address one of the most critical impediments intrinsic in MRI, which is the relatively slow speed of image acquisition. Although very significant, such an improvement may still be insufficient in the cases when a repetitive acquisition of MRI scans pertaining to the same volume of interest is required. Acquisitions of this type are prevalent in diffusion MRI, in which an independent MRI scan is required to encode the strength of water diffusion along a predefined spatial direction. Thus, for example, an accurate delineation of multimodal diffusion profiles by means of high angular resolution diffusion imaging (HARDI) requires using between 60 and 100 diffusion-encoding gradients. This, in turn, is translated into relatively long acquisition times, which adversely affects the applicability of HARDI for clinical diagnosing. To overcome this limitation, the present paper introduces a method for substantial reduction of the number of diffusion encoding gradients required for reliable reconstruction of HARDI signals. The method exploits the theory of CS, which establishes conditions on which a signal of interest can be recovered from its under-sampled measurements, provided that the signal admits a sparse representation in the domain of a linear transform. In the case at hand, the latter is defined to be spherical ridgelet transformation, which excels in sparsifying HARDI

*O. Michailovich and S. Dolui are with the School of Electrical and Computer Engineering, University of Waterloo, Canada N2L 3G1 (e-mails: {olegm,sdolui}@uwaterloo.ca). Y. Rathi is with the Psychiatry Neuroimaging Laboratory (Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School), Boston, MA 02115 USA, (e-mail: yogesh@bwh.harvard.edu.)

signals. What makes the resulting reconstruction procedure even more accurate is a combination of the sparsity constraints in the diffusion domain with additional constraints imposed on the estimated diffusion field in the spatial domain. Accordingly, the present paper describes a novel way to combine the diffusion- and spatial-domain constraints to achieve a maximal reduction in the number of diffusion measurements, while sacrificing little in terms of reconstruction accuracy. Finally, details are provided on a particularly efficient numerical scheme which can be used to solve the aforementioned reconstruction problem by means of standard and readily available estimation tools. The paper is concluded with experimental results which support the practical value of the proposed reconstruction methodology.

1 Introduction

The human brain consists of about 10^{11} nerve cells that can be subdivided into about 1000 different cell types, a complexity that far exceeds that of other organs of the body. A further complexity is evident in the way in which the component cells of the brain interconnect and function [1]. In contrast to other types of the cells, each neuron communicates with many target cells by means of its distinctive protoplasmatic protrusion, called an axon. Axons with similar destinations, in turn, tend to form bundles - known as neural fibre tracts - which play a pivotal role in the determination of brain connectivity. Through reconstructing the pattern of connectivity of the neural tracts in both healthy and diseased subjects, it is therefore possible to obtain an abundance of valuable diagnostic information that could be used for early diagnostics of brain-related disorders, for assessing the damage caused to the brain by stroke, tumours or injuries, as well as for planning and monitoring of neurosurgeries [2].

Central to MRI is the notion of contrast, which is typically defined by the biochemical composition of interrogated tissue as well as by the morphology of its associated parenchyma. Prevalent in MRI practice are the contrasts determined by the T_1/T_2 relaxation times and proton density (PD). Despite their exceptional importance to clinical diagnosis, none of the above contrast mechanisms has demonstrated effectiveness in delineating the morphological structure of the white matter. It is only with the advent of diffusion MRI (dMRI) that scientists have been able to perform quantitative measurements of the diffusivity of white matter, based on which its structural delineation has become possible [2–8]. Although over the last two decades dMRI has developed into an established technique with a great impact on health care and neurosciences, like any other MRI technique it remains subject to artifacts and pitfalls [3]. While many of such artifacts can be overcome by using advanced hardware designs and/or more sophisticated imaging protocols [2, Ch.2], [9], one particularly critical limitation of dMRI stems from the physics of the acquisition of diffusion MR images, and therefore is impossible to resolve by operational means. Specifically, since collecting the dMRI data requires a repetitive acquisition of MR responses from the same volume for a number of diffusion-encoding gradi-

ents, it is the relatively long acquisition times that greatly impair the practical value of this important imaging modality. Particularly, longer acquisition times entail a higher probability for the patient to exercise involuntary motion (typically caused by fatigue and/or stress related tremors, swallowing, uncontrollable sighing or coughing), which severely affects the quality of dMRI data. Moreover, since during the whole duration of the scan the patients are required to remain motionless, it is currently deemed ineffective to apply dMRI-based diagnosis in paediatrics as well as to patients with dementia or post-traumatic injuries, where non-compliance is typical. The problem of long acquisition times also hampers the application of dMRI for intra-operational imaging, where it could be an irreplaceable tool to use for neurosurgical planning and decision-making support [10–12]. Lastly, relatively long scanning times required by dMRI aggravate the problem of accessibility to MRI equipment. All the above arguments suggest that the practical value of dMRI could be improved by shortening the scanning times required for acquisition of dMRI data. A particular method to achieve such an improvement is detailed in this paper.

In this work, we adopt a general diffusion model in which each voxel within a region of interest (ROI) is allowed to support more than one fibre tract. In this case, under some general assumptions (see, e.g., [13, Sect. 3.1] for more details), the diffusion signal $s(\mathbf{u} | \mathbf{r})$ originating from a voxel with spatial coordinate $\mathbf{r} \in \mathbb{R}^3$ containing $M(\mathbf{r})$ fibres can be modelled as [13, 14]

$$s(\mathbf{u} | \mathbf{r}) = s_0(\mathbf{r}) \sum_{i=1}^{M(\mathbf{r})} \alpha_i(\mathbf{r}) \exp \left\{ -b(\mathbf{u}^T D_i(\mathbf{r}) \mathbf{u}) \right\}, \quad (1)$$

where \mathbf{u} denotes the spherical coordinate, i.e.

$$\mathbf{u} \in \mathbb{S}^2 := \{ \mathbf{v} \in \mathbb{R}^3 \mid \|\mathbf{v}\|_2 = 1 \}, \quad (2)$$

and $\alpha_i(\mathbf{r}) > 0$ are positive weights obeying $\sum_{i=1}^{M(\mathbf{r})} \alpha_i(\mathbf{r}) = 1$. In (1), s_0 denotes the diffusion signal obtained in the absence of diffusion encoding (i.e. the so-called “b0 image”), b is defined as a function of the shape and amplitude of diffusion-encoding gradients [15, Eq. 3.18], and $\{D_i(\mathbf{r})\}_{i=1}^{M(\mathbf{r})}$ are 3×3 diffusion tensors associated with the $M(\mathbf{r})$ neural fibres passing through the coordinate \mathbf{r} . In practical settings, the spherical coordinate \mathbf{u} is sampled at K distinct points $\{\mathbf{u}_k\}_{k=1}^K$ over the unit sphere. In this case, for each \mathbf{u}_k , MR measurements are acquired in the form of a diffusion-encoded image $s_k(\mathbf{r}) := s(\mathbf{u}_k | \mathbf{r})$. As a result, a typical dMRI data set consists of a collection of such diffusion-encoded images $\{s_k(\mathbf{r})\}_{k=1}^K$, whose size K determines the accuracy with which the directions of local diffusion flows can be estimated.

Provided unlimited scanning time, one could measure the diffusion in thousands of orientations, making it possible to identify the directions of dominant diffusion with very high precision. For the reasons explained earlier, however, scanning times are always limited, which necessitates restriction of K to a reasonably small value. This brings us to the central question addressed in the present paper: what is a sufficient number K of diffusion-encoding directions to

use? It turns out that, for some realizations of dMRI, the above question can be answered in a rigorous manner. In particular, in diffusion tensor imaging (DTI) [3–5, 7, 8, 16, 17], the reconstruction is carried out under the assumption that each voxel can support only one diffusion flow as most, which corresponds to setting $M(\mathbf{r}) = 1$ for all \mathbf{r} in (1). Accordingly, a minimum of $K = 7$ diffusion-encoded images are theoretically sufficient to measure $S_0(\mathbf{r})$ and recover the six non-repetitive components of the symmetric tensors $D(\mathbf{r})$ by means of least-square fitting. (In practice, however, a larger number of gradient directions is employed to reduce the estimation variance, with a typical K being between 25 and 30 [18].) Unfortunately, the accuracy of DTI is known to deteriorate dramatically at the sites where the neural fibres (or bundles thereof) cross, touch upon each other, or diverge [13, 19–27].

The fibre crossing problem in DTI has prompted efforts to develop dMRI methodologies which are capable of detecting multiple diffusion flows (or, equivalently, neural fibre tracts) within a given voxel. One of such techniques is High Angular Resolution Diffusion Imaging (HARDI) [14, 20, 22–24, 26–28], which is capable of capturing multi-modal diffusion patterns by sampling the spherical shell at a much greater number of points (usually between 60 and 100) as compared to the case of DTI. Increasing K makes it possible to describe the diffusion measurements using much more accurate models. Among these are parametric models [29–34] which allow HARDI signals to be expressed in terms of a relatively small number of prototype signal forms. Unfortunately, fitting a parametric model often entails minimization of non-convex functionals, which is a noise-sensitive and computationally intensive task, prone to the problem of local minima. The need to predetermine the optimal number of fitting terms is known to be another disadvantage of using the models of the above type.

The problems associated with parametric modelling of HARDI signals can be overcome by using non-parametric models, in which case the signals are recovered by projecting the observed data onto properly defined functional subspaces. In particular, the applicability of spherical Fourier analysis [35] to dMRI has been demonstrated in [25–28], where HARDI signals are approximated by truncated series of spherical harmonics (SH). Despite its stability and computational efficiency, however, the SH-parameterization involves a relatively large number of SHs, which suggests that the SHs cannot be an adequate basis for sparse representation of HARDI signals. The main reason for this is rooted in the fact that the energies of elementary signals $d_i(\mathbf{u}|\mathbf{r}) := \exp\{-b(\mathbf{u}^T D_i(\mathbf{r})\mathbf{u})\}$ in (1) are concentrated alongside the great circles of \mathbb{S}^2 , whereas the energy of SHs is spread all over \mathbb{S}^2 , and, as a result, a relatively large number of SHs are needed to effectively “encode” each of d_i . The inability of the basis of SHs to sparsely represent diffusion signals has led to the proposal of spherical ridgelets in [36, 37], where it was shown that it only takes 6 to 8 spherical ridgelets on average to represent the HARDI signals with a precision exceeding the precision of their representation with 45 SHs.

The present work takes the ideas of [36, 37] one step further and shows that the availability of a sparsifying basis for HARDI signals can be used to reduce the number of diffusion gradients required for data acquisition. In particular,

we suggest to use the theory of compressed sensing (CS) [38–44] to recover the HARDI signals using the number of spherical samples K in a range of values typical for DTI (i.e. $K \in [16, 24]$), thus allowing a multi-fibre analysis of dMRI data to be performed at the “acquisition cost” of a standard DTI.

It is worthwhile noting that the ideas of CS have already paved their way into the field of diffusion imaging [45–49]. In this regard, conceptually the closest to the proposed approach is the method reported in [46]. In spite of this similarity, however, there are two principal distinctions which make the present method a more powerful alternative. In particular, the basis functions used in [46] are limited to represent an *average* diffusivity and anisotropy of the white matter, thereby neglecting both intra- and inter-voxel variability of tensors $D_i(\mathbf{r})$ in (1). The ridgelet representation, on the other hand, is a multiresolution technique, which possesses an intrinsic ability to deal with a continuum of different diffusion scales. Second, the approach in [46] is applied in a “voxel-by-voxel” manner and it therefore does not take into consideration the spatial regularity of diffusion field. The present paper, on the other hand, proposes a novel formulation of the problem of CS-based reconstruction of diffusion signals, in which the sparsity constraints enforced in the diffusion domain are augmented by regularity constraints enforced in the spatial domain. The resulting reconstruction problem has the format of a convex minimization problem, which is solved using a specially adapted version of the split Bregman algorithm [50–52]. As will be shown below, the proposed algorithm results in a particularly advantageous computational structure which allows the solution to be computed via a sequence of simple and easily parallelizable steps.

The rest of the paper is organized as follows. Section II provides additional comments on the input-output structure of the proposed algorithm. The construction of spherical ridgelets is briefly outlined in Section III, whereas Section IV gives a formal description of the proposed reconstruction methodology. Some principal details on the numerical implementations of the proposed algorithm are summarized in Section V, with the results of our experimental studies reported in Section VI. Section VII finalizes the paper with a discussion and conclusions.

2 Problem Statement

In the centre of our considerations is the diffusion signal $s(\mathbf{u}|\mathbf{r})$ which, when normalized by its related $b0$ -image $s_0(\mathbf{r})$, quantifies the attenuation of MR read-out caused by the diffusion of water molecules in the direction $\mathbf{u} \in \mathbb{S}^2$ through the spatial position $\mathbf{r} \in \mathbb{R}^3$. In practical settings, both \mathbf{u} and \mathbf{r} are discretized. Specifically, restricting \mathbf{u} to a discrete set of orientations $\{\mathbf{u}_k\}_{k=1}^K$ prescribes the acquisition of diffusion data in the form of K diffusion-encoded images $\{s_k(\mathbf{r})\}_{k=1}^K$, with each $s_k(\mathbf{r})$ corresponding to a given \mathbf{u}_k . In this case, for a fixed \mathbf{r}_0 , the vector $[s_1(\mathbf{r}_0), s_2(\mathbf{r}_0), \dots, s_K(\mathbf{r}_0)]^T \in \mathbb{R}^K$ represents a discretization of $s(\mathbf{u}|\mathbf{r}_0)$. Note that such a discretization follows a linear measurement model, since each sample $s_k(\mathbf{r}_0)$ can be represented as an inner product of $s(\mathbf{u}|\mathbf{r}_0)$ with a sampling function. In particular, let $\{\varphi_k(\mathbf{u})\}_{k=1}^K$ be a Dirac basis of sampling

functions defined by

$$\varphi_k(\mathbf{u}) = \delta(1 - \mathbf{u} \cdot \mathbf{u}_k), \quad k = 1, 2, \dots, K, \quad (3)$$

where δ denotes the Dirac delta function and the dot stands for the Euclidean dot product. Then, formally,

$$s_k(\mathbf{r}_0) = \langle s(\cdot | \mathbf{r}_0), \varphi_k \rangle_{\mathbb{L}_2} := \int_{\mathbb{S}^2} s(\mathbf{u} | \mathbf{r}_0) \varphi_k(\mathbf{u}) d\eta(\mathbf{u}), \quad \text{with } k = 1, 2, \dots, K, \quad (4)$$

with $d\eta$ being the standard rotation invariant measure on \mathbb{S}^2 .

Next, given a collection of M spherical ridgelets $\{\psi_m(\mathbf{u})\}_{m=1}^M$ (defined below), the signal $s(\mathbf{u} | \mathbf{r})$ is assumed to be expandable as

$$s(\mathbf{u} | \mathbf{r}) = \sum_{m=1}^M c(\mathbf{r}) \psi_m(\mathbf{u}), \quad (5)$$

with $c(\mathbf{r}) \in \mathbb{R}^M$ being a vector of spherical ridgelet coefficients which depend on the spatial coordinate \mathbf{r} . It is important to note that the set of spherical ridgelets is allowed to be overcomplete, implying $\dim[\text{Span}\{\psi_m(\mathbf{u})\}_{m=1}^M] < M$. A practical consequence of this fact is that the definition of coefficients $c(\mathbf{r})$ in (5) is, in general, not unique. This non-uniqueness is further aggravated by the fact that $c(\mathbf{r})$ will have to be recovered from an under-sampled set of diffusion measurements, in which case $K \ll M$. Overcoming such a severe underdetermination in the problem of estimating the ridgelet coefficients $c(\mathbf{r})$ will be possible based on the fundamental premise of the theory of CS, which states that an accurate estimation of $c(\mathbf{r})$ is possible if the latter is sufficiently sparse *and* if the sampling and representation bases are sufficiently decorrelated. While the sparsity of $c(\mathbf{r})$ is rooted in the very design of spherical ridgelets [37], the incoherency between the Dirac sampling functions (3) and spherical ridgelets stems from the fact that the former have an infinitely small support, whereas the latter are “smeared” all over the unit sphere. The above properties of the basis of spherical ridgelets yield conditions for an effective application of CS, in which case one can obtain a faithful reconstruction of diffusion signals using as few as $K = 16$ diffusion-encoding gradients.

The proposed algorithm produces an estimate of the ridgelet representation coefficients $c(\mathbf{r})$ in (5). Once available, the coefficients provide an access to the analytical definition of diffusion signals by virtue of (5). This can be used in a number of ways. One possibility could be to use the ridgelet coefficients to compute their associated orientation distribution functions [14], based on which a multi-fibre tractography analysis can be done [53,54]. Alternatively, (5) can be used to evaluate the diffusion signals over an arbitrarily fine grid of orientations. Subsequently, such refined “measurements” could be fitted using a different representation model, whose application to the original data would not have been possible without causing severe underestimation errors. Deconvolving the refined data to estimate the underlying fibre orientation functions [55–58] would be another important option to follow. In this paper, we refrain from questioning

which of the above possibilities is more advantageous over the others. Our sole objective here is to specify a signal processing algorithm which can be used to recover HARDI signals, while using the number of diffusion-encoded images typical for a standard DTI.

Finally, it should be noted that the primarily purpose of the proposed methodology is to improve the value of HARDI in terms of its time efficiency. Since the improvement is achieved through merely decreasing the number of diffusion-encoding gradients, the proposed method by no means abrogates the use of fast imaging protocols [59–61] to further accelerate the data acquisition. Furthermore, an additional speed-up can be achieved via applying CS to reconstruct the diffusion encoded images $s_k(\mathbf{r})$ from their subcritical samples in the spectral domain [62–66]. Generally speaking, we believe this is a combination of such software and hardware technologies which will eventually lead to substantial improvements in the practical value of HARDI-based diagnosing. In this paper, however, we confine our contribution to showing one particular way of attaining this important objective.

3 Spherical ridgelets

It is the property of spherical ridgelets to provide sparse representation of diffusion signals described by (1) which makes them an unparalleled tool for CS-reconstruction of HARDI data. To avoid repetitions, in what follows, we present only the most principal points of ridgelets design, while their detailed description can be found in [37].

Spherical ridgelets are constructed using the fundamental principles of wavelet theory [67, 68]. Specifically, let $x \in \mathbb{R}_+$ and $\rho \in (0, 1)$ be a positive scaling parameter. Further, let $\kappa(x) = \exp\{-\rho x(x+1)\}$ be a Gaussian function, which we subject to a series of dyadic scalings to result in

$$\kappa_j(x) = \kappa(2^{-j}x) = \exp\left\{-\rho \frac{x}{2^j} \left(\frac{x}{2^j} + 1\right)\right\}, \quad (6)$$

where $j \in \mathbb{N} := \{0, 1, 2, \dots\}$. Subsequently, the Gaussian-Weierstrass scaling function $\chi_{j,\mathbf{v}} : \mathbb{S}^2 \rightarrow \mathbb{R}$ at resolution $j \in \mathbb{N}$ and orientation $\mathbf{v} \in \mathbb{S}^2$ can be defined as given by [69, 70]

$$\chi_{j,\mathbf{v}}(\mathbf{u}) = \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \kappa_j(n) P_n(\mathbf{u} \cdot \mathbf{v}), \quad \forall \mathbf{u} \in \mathbb{S}^2, \quad (7)$$

where P_n denotes the Legendre polynomial of order n . It is worth noting that the \mathbb{L}_2 energy of $\chi_{j,\mathbf{v}}$ is concentrated around the spherical point \mathbf{v} , with this concentration becoming more and more localized when j approaches infinity.

The spherical ridgelets are designed with the help of the Funk-Radon transform which, for an arbitrary continuous function $f : \mathbb{S}^2 \rightarrow \mathbb{R}$, is defined as

$$\mathcal{R}\{f\}(\mathbf{v}) = \int_{\mathbf{u} \in \sigma(\mathbf{v})} f(\mathbf{u}) \eta(\mathbf{u}), \quad (8)$$

with $\sigma(\mathbf{v})$ denoting the great circle perpendicular to direction \mathbf{v} , i.e. $\sigma(\mathbf{v}) := \{\mathbf{u} \in \mathbb{S}^2 \mid \mathbf{u} \cdot \mathbf{v} = 0\}$. Subsequently, following [37], the semi-discrete frame \mathbb{U} of spherical ridgelets can be defined as

$$\mathbb{U} := \{\psi_{j,\mathbf{v}} \mid \mathbf{v} \in \mathbb{S}^2, j = -1, 0, 1, 2, \dots\}, \quad (9)$$

where the spherical ridgelet functions $\psi_{j,\mathbf{v}}$ are obtained from $\chi_{j,\mathbf{v}}$ according to

$$\psi_{j,\mathbf{v}}(\mathbf{u}) = \frac{1}{2\pi} \mathcal{R}\{\chi_{j+1,\mathbf{v}} - \chi_{j,\mathbf{v}}\}(\mathbf{u}), \quad (10)$$

with $\chi_{-1,\mathbf{v}}(\mathbf{u}) \equiv 0$. Using (7), the ridgelets (10) can be redefined in a closed form as (see [35] for details)

$$\psi_{j,\mathbf{v}}(\mathbf{u}) = \frac{1}{2\pi} \sum_{n=0}^{\infty} \frac{2n+1}{4\pi} \lambda_n (\kappa_{j+1}(n) - \kappa_j(n)) P_n(\mathbf{u} \cdot \mathbf{v}), \quad (11)$$

where $\kappa_{-1}(n) = 0, \forall n$ and

$$\lambda_n = \begin{cases} 2\pi(-1)^{n/2} \frac{1 \cdot 3 \cdots (n-1)}{2 \cdot 4 \cdots n}, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd.} \end{cases} \quad (12)$$

The set \mathbb{U} in (9) is infinite-dimensional, and hence is not suitable for practical computations. To define a discrete counterpart of \mathbb{U} , one has first to restrict the values of the resolution index j to a finite set $\{-1, 0, 1, \dots, J\}$, where J defines the highest level of “detectable” signal details. Additionally, the set of all possible *orientations* $\mathbf{v} \in \mathbb{S}^2$ of spherical ridgelets needs to be discretized as well. To find a proper discretization scheme, we first note that the construction in (11) suggests that the bandwidth of the spherical ridgelets (and therefore the dimensionality of the functional space they belong to) increases proportionally to 2^j . Since the space of spherical harmonics of degree n has a dimension of $(n+1)^2$, it seems to be reasonable to define the number of ridgelet orientations at resolution j to be equal to $M_j = (2^{j+1}m_0 + 1)^2$, with m_0 being the smallest spherical order resulting in $\kappa_0(m_0) \leq \epsilon$ for some predefined $0 < \epsilon \ll 1$ (e.g. $\epsilon = 10^{-6}$). Consequently, for each j , a total of M_j orientations $\{\mathbf{v}_j^i\}_{i=1}^{M_j}$ are chosen so that a discrete counterpart of \mathbb{U} can now be defined as

$$\mathbb{U}_d = \left\{ \psi_{j,\mathbf{v}_j^i} \mid j = -1, 0, 1, \dots, J, i = 1, 2, \dots, M_j \right\}. \quad (13)$$

where the subscript d stands for “discrete”. It should be noted that, although the set \mathbb{U}_d is composed of continuously defined functions, its dimension is finite, since \mathbb{U}_d consists of a total of $M = \sum_{j=-1}^J (2^{j+1}m_0 + 1)^2$ spherical ridgelets. To slightly simplify our notation, in what follows, the spherical ridgelets in \mathbb{U}_d will be indexed as $\psi_m(\mathbf{u})$, with $m = 1, 2, \dots, M$ being a combined index accounting for both different resolutions and orientations.

Given a sampling set of K diffusion-encoding orientations $\{\mathbf{u}_k\}_{k=1}^K$, one can use (11) to compute the values of the spherical ridgelets in \mathbb{U}_d over the sampling

set¹. The resulting values can be stored into a $K \times M$ matrix A defined as

$$A = \begin{bmatrix} \psi_1(\mathbf{u}_1) & \psi_2(\mathbf{u}_1) & \dots & \psi_M(\mathbf{u}_1) \\ \psi_1(\mathbf{u}_2) & \psi_2(\mathbf{u}_2) & \dots & \psi_M(\mathbf{u}_2) \\ \dots & \dots & \dots & \dots \\ \psi_1(\mathbf{u}_K) & \psi_2(\mathbf{u}_K) & \dots & \psi_M(\mathbf{u}_K) \end{bmatrix}. \quad (14)$$

Then, for a given vector $s(\mathbf{r}) := [s(\mathbf{u}_1 | \mathbf{r}), s(\mathbf{u}_2 | \mathbf{r}), \dots, s(\mathbf{u}_K | \mathbf{r})]^T \in \mathbb{R}^K$ of the measured values of a diffusion signal $s(\mathbf{u} | \mathbf{r})$ at the spatial location \mathbf{r} , the model (5) asserts the existence of representation coefficients $c(\mathbf{r}) \in \mathbb{R}^M$ such that

$$s(\mathbf{r}) = A c(\mathbf{r}) + e(\mathbf{r}), \quad (15)$$

where $e(\mathbf{r})$ accounts for both model and measurement noises. Clearly, the non-negligibility of $e(\mathbf{r})$ along with the fact that $K \ll M$ makes the problem of recovering the representation coefficients $c(\mathbf{r})$ from $s(\mathbf{r})$ a very challenging inverse problem, our solution to which is presented next.

4 Proposed reconstruction framework

Let Ω represent the volume within which diffusion measurements are acquired. Also known as a region of interest, Ω is assumed to be a bounded rectangular subdomain of \mathbb{R}^3 , i.e. $\Omega := [0, L_x] \times [0, L_y] \times [0, L_z] \subset \mathbb{R}^3$. Let Ω_d be a discrete subset of Ω , which represents the spatial locations at which the diffusion signal is measured. Specifically, Ω_d is assumed to be a uniform lattice which can be formally defined as

$$\Omega_d := \left\{ \mathbf{r} = \{x_{i_1}, y_{i_2}, z_{i_3}\} \mid x_{i_1} = \frac{i_1}{N_x} L_x, y_{i_2} = \frac{i_2}{N_y} L_y, z_{i_3} = \frac{i_3}{N_z} L_z \right\}, \quad (16)$$

where $0 \leq i_1 < N_x$, $0 \leq i_2 < N_y$, and $0 \leq i_3 < N_z$ are sampling indices in the direction of x , y and z coordinates, respectively.

Let K be the number of diffusion-encoding gradients used for HARDI data acquisition, and let the corresponding gradient orientations be denoted by $\{\mathbf{u}_k\}_{k=1}^K$, where $\mathbf{u}_k \in \mathbb{S}^2$. For each of these values of \mathbf{u}_k , MRI measurements result in its associated diffusion-encoded image s_k , which can be formally viewed as a mapping from Ω_d to \mathbb{R} . For the sake of convenience, each s_k can be stored and manipulated as an $N_x \times N_y \times N_z$ array of real numbers, namely $s_k \in \mathbb{R}^{N_x \times N_y \times N_z}$. Alternatively, at a given coordinate $\mathbf{r} \in \Omega_d$, one can combine the values $s_1(\mathbf{r}), s_2(\mathbf{r}), \dots, s_K(\mathbf{r})$ into a column vector $s(\mathbf{r}) := [s_1(\mathbf{r}), s_2(\mathbf{r}), \dots, s_K(\mathbf{r})]^T \in \mathbb{R}^K$ (as it was already done in (15)). This vector can then be regarded as a vector of discrete measurements of an associated HARDI signal $s(\mathbf{u} | \mathbf{r})$ measured at orientations $\{\mathbf{u}_k\}_{k=1}^K$. It is worth noting that, according to the above notations,

¹Since the definition in (11) involves an infinite summation, the latter needs to be truncated to render the computations practical. In practice, we truncate the summation to index n_{max} for which the magnitude of the summand drops below 10^{-9} .

the value $s_k(\mathbf{r})$ admits a twofold interpretation, *viz.* either as the k^{th} coordinate of vector $s(\mathbf{r})$ or as the value of image s_k at spatial position \mathbf{r} .

When combined together, the continuum of vectors $s(\mathbf{r})$ can be regarded as a discrete vector field $s : \Omega \rightarrow \mathbb{R}^K$, in which case $s(\mathbf{r})$ has a natural interpretation of the value of s corresponding to position \mathbf{r} . The vector space \mathfrak{V} of such vector fields can be endowed with the standard inner product

$$\langle s^1, s^2 \rangle_{\mathfrak{V}} = \sum_{\mathbf{r} \in \Omega_d} s^1(\mathbf{r})^T s^2(\mathbf{r}) = \sum_{\mathbf{r} \in \Omega_d} \sum_{k=1}^K s_k^1(\mathbf{r}) s_k^2(\mathbf{r}) = \sum_{k=1}^K \langle s_k^1, s_k^2 \rangle, \quad (17)$$

with $\langle s_k^1, s_k^2 \rangle = \sum_{\mathbf{r} \in \Omega_d} s_k^1(\mathbf{r}) s_k^2(\mathbf{r})$ being the standard inner product between the scalar-valued images s_k^1 and s_k^2 . Accordingly, congruent to the definition in (17), the ℓ_2 -norm of $s \in \mathfrak{V}$ is defined as

$$\|s\|_{\mathfrak{V},2} = \left[\sum_{\mathbf{r} \in \Omega_d} \|s(\mathbf{r})\|_2^2 \right]^{1/2} = \left[\sum_{k=1}^K \|s_k\|_F^2 \right]^{1/2}, \quad (18)$$

where $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the Euclidean vector and Frobenius matrix norms, respectively.

Another norm on \mathfrak{V} that we shall make use of is the total variation (TV) semi-norm which is defined as follows. First, let us define the total variation of the k^{th} component s_k of the field s in a standard manner as

$$\|s_k\|_{\text{TV}} = \sum_{\mathbf{r} \in \Omega_d} \left[\sum_{\mathbf{p} \in \mathcal{C}(\mathbf{r})} |s_k(\mathbf{r}) - s_k(\mathbf{p})|^2 \right]^{1/2}, \quad (19)$$

where $\mathcal{C}(\mathbf{r} = (x_{i_1}, y_{i_2}, z_{i_3})) = \{(x_{i_1-1}, y_{i_2}, z_{i_3}), (x_{i_1}, y_{i_2-1}, z_{i_3}), (x_{i_1}, y_{i_2}, z_{i_3-1})\}$ is a 3-neighbourhood (causal) clique of voxel \mathbf{r} . Consequently, the TV norm of the discrete vector field s can be defined in terms of the TV-norms of its K components as

$$\|s\|_{\mathfrak{V},\text{TV}} = \left[\sum_{k=1}^K \|s_k\|_{\text{TV}}^\alpha \right]^{1/\alpha}. \quad (20)$$

Thus, for example, $\alpha = 2$ was used in the TV-denoising method reported in [71]. In this paper, we use $\alpha = 1$.

Now, let $\{\psi_m\}_{m=1}^M$ be a set of spherical ridgelets defined by (13), which is assumed to be rich enough so that each HARDI signal can be expressed according to (5). Analogously to the discrete measurements $s(\mathbf{r})$, the representation coefficients $c(\mathbf{r})$ corresponding to different voxels \mathbf{r} can be aggregated into a vector field $c \in \mathfrak{U}$, where $\mathfrak{U} : \Omega_d \rightarrow \mathbb{R}^M$ (with $c(\mathbf{r})$ being the value of c observed at \mathbf{r}). Although it is possible to endow the vector space \mathfrak{U} with both the ℓ_2 - and TV-norms by analogy with (18) and (20), it will be particularly useful to consider the ℓ_1 -norm of c which can be defined as

$$\|c\|_{\mathfrak{U},1} = \sum_{\mathbf{r} \in \Omega_d} \|c(\mathbf{r})\|_1 = \sum_{\mathbf{r} \in \Omega_d} \sum_{k=1}^M |c_k(\mathbf{r})|. \quad (21)$$

Using the definitions of the vector fields \mathfrak{V} and \mathfrak{U} as well as the definition of A in (14), a connection between \mathfrak{V} and \mathfrak{U} is established by means of a linear map $\mathcal{A} : \mathfrak{U} \rightarrow \mathfrak{V}$ that is defined as given by

$$\mathcal{A} : \mathfrak{U} \rightarrow \mathfrak{V} : c(\mathbf{r}) \mapsto s(\mathbf{r}) = A c(\mathbf{r}), \forall \mathbf{r} \in \Omega_d. \quad (22)$$

Consequently, using \mathcal{A} , one can define the HARDI data formation model as

$$s = \mathcal{A}\{c\} + e, \quad (23)$$

where $e \in \mathfrak{V}$ is supposed to account for both measurement noise and modelling errors, and it is assumed to have a relatively small ℓ_2 -norm $\|e\|_{\mathfrak{V},2} \leq \epsilon$.

The model (23) suggests a reduction of the problem of estimation of HARDI signals to the problem of estimation of their corresponding representation coefficients c from the discrete and noisy measurements s . Furthermore, as our main intension has been to recover the coefficients c using as few diffusion-encoding gradients as possible (implying $K \ll M$), there is an infinite number of solutions which would fit the constraint $\|\mathcal{A}\{c\} - s\|_{\mathfrak{V},2} \leq \epsilon$. However, if it is known *a priori* that, for each $\mathbf{r} \in \Omega_d$, the vector of representation coefficients $c(\mathbf{r})$ is sparse, then a useful estimate of c can be obtained as a solution to the following convex optimization problem [38–44]

$$\min_c \|c\|_{\mathfrak{V},1} \quad (24)$$

$$\text{s.t. } \|\mathcal{A}\{c\} - s\|_{\mathfrak{V},2} \leq \epsilon. \quad (25)$$

It should be noted that the optimization problem (24) is equivalent to solving

$$\min_c \sum_{\mathbf{r} \in \Omega_d} \|c(\mathbf{r})\|_1 \quad (26)$$

$$\text{s.t. } \sum_{\mathbf{r} \in \Omega_d} \|A c(\mathbf{r}) - s(\mathbf{r})\|_2^2 \leq \epsilon^2, \quad (27)$$

and therefore, under the assumption of spatially homogeneous noise e , the problem (24) is separable in the spatial coordinate \mathbf{r} . This means that an optimal field c can be recovered by solving for its components

$$\min_{c(\mathbf{r})} \|c(\mathbf{r})\|_1 \quad (28)$$

$$\text{s.t. } \|\mathcal{A}\{c(\mathbf{r})\} - s(\mathbf{r})\|_2 \leq (N_x N_y N_z)^{-1/2} \epsilon, \quad (29)$$

independently at each $\mathbf{r} \in \Omega_d$.

While computationally attractive, the above solution is suboptimal, since it completely disregards the dependencies which are likely to exist between spatially adjacent HARDI signals. A possible way to take such dependencies into consideration is to require the noise-free version of the measured signal s to possess a minimal TV norm among all possible candidate solutions [72]. This

requirement can be translated into the following minimization problem

$$\begin{aligned} \min_c \left\{ \|c\|_{\mathfrak{V},1} + \gamma \|\mathcal{A}\{c\}\|_{\mathfrak{V},\text{TV}} \right\} \\ \text{s.t. } \|\mathcal{A}\{c\} - s\|_{\mathfrak{V},2} \leq \epsilon. \end{aligned} \quad (30)$$

where the role of $\gamma > 0$ is to balance the relative influence of the sparse and TV terms in the above cost function. The optimization problem (30) can be rewritten in its equivalent Lagrangian form

$$\min_c \left\{ \frac{1}{2} \|\mathcal{A}\{c\} - s\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} + \mu \|\mathcal{A}\{c\}\|_{\mathfrak{V},\text{TV}} \right\}, \quad (31)$$

for some optimal values of $\lambda > 0$ and $\mu > 0$ [73].

Below, we are going to specify a particular, computationally efficient method for solving (31). In this connection, it is instructive to outline the following two instances of (31).

4.1 Sparse-only reconstruction

When $\mu = 0$, the functional in (31) becomes *separable in the spatial variable* \mathbf{r} in the sense that, in such a case, an optimal c can be recovered by solving

$$\min_{c(\mathbf{r})} \left\{ \frac{1}{2} \|A c(\mathbf{r}) - s(\mathbf{r})\|_2^2 + \lambda \|c(\mathbf{r})\|_1 \right\} \quad (32)$$

for each $c(\mathbf{r})$ independently. Note that (32) can be considered to be a Lagrangian form of the optimization problem (28). There exist a broad spectrum of methods which could be used for solution of (32). Some particularly attractive algorithms seem to be those exploiting the principle of iterative shrinkage (aka iterated thresholding) [74–77]. While the non-differentiability of ℓ_1 -norm in (32) rules out the applicability of gradient-based optimization tools, iterative shrinkage methods are capable of finding a solution of (32) through iterative application of a first-order, fixed-point update rule. Specifically, many algorithms of this kind find an optimal solution as a stationary point of the sequence of estimates produced by

$$c(\mathbf{r})^{t+1} = \mathcal{S}_{\lambda/\nu} \left\{ c(\mathbf{r})^t + \nu^{-1} A^T (s(\mathbf{r}) - A c(\mathbf{r})^t) \right\}, \quad (33)$$

where $\mathcal{S}_\tau\{t\} = \text{sign}(t)(|t| - \tau)_+$ denotes the operator of soft thresholding and ν is chosen to obey $\nu > \|AA^T\|$. In the present paper, a modification of the iterative update in (33), known as the *fast iterative shrinkage thresholding algorithm* (FISTA) [78], was employed due to its considerably faster convergence as compared with many alternative “accelerated” methods.

It should be emphasized that, while being suboptimal from the viewpoint of spatial-domain regularity, the solution of (32) through iterative shrinkage is advantageous in two important practical ways. First, it suggests considerable

storage reduction, since the thresholding operator in (33) sets to zero the representation coefficients with amplitudes less or equal to λ/ν in absolute value. It makes it possible to use sparse data formats to store and manipulate the representation coefficients. Second, the fact that the estimation of $c(\mathbf{r})$ is performed at each voxel independently suggests a natural way to speed up the overall estimation process though parallel computing on a multicore system.

4.2 TV-only reconstruction

When $\lambda = 0$, solving the optimization problem (31) is equivalent to simultaneously solving K optimization problems of the form

$$\min_c \left\{ \frac{1}{2} \|[\mathcal{A}\{c\}]_k - s_k\|_F^2 + \mu \|[\mathcal{A}\{c\}]_k\|_{\text{TV}} \right\}, \quad (34)$$

where $k = 0, 1, \dots, K-1$ and $[\mathcal{A}\{c\}]_k$ denotes the k -th component of the vector field $\mathcal{A}\{c\} \in \mathfrak{V}$. Let $[\mathcal{A}\{c\}]_k$ be denoted by u_k , i.e. $u_k := [\mathcal{A}\{c\}]_k$. Then, reformulated with respect to u_k , the problem (34) can be rewritten as

$$\min_{u_k} \left\{ \frac{1}{2} \|u_k - s_k\|_F^2 + \mu \|u_k\|_{\text{TV}} \right\}, \quad (35)$$

in which case it can be recognized as the problem of TV-denoising of the diffusion-encoded image s_k [72]. It is important to note that, in contrast to (34), the problem (35) can be solved for each k independently, in which case we say that the estimation becomes *separable in the diffusion direction*.

The current arsenal of methods which can be used for solving (35) is broad. Originated from the work of Rudin *et al* [72], TV-denoising methods currently include gradient-based optimization methods [79], first-order methods [80], iterative shrinkage methods [81,82], and Bregman-type iterative algorithms [50,51,83]. The first-order methods appear to be a particularly attractive option when relatively large data arrays need to be processed, which is the case relevant to dMRI. Consequently, in the present paper, we employ the algorithm of [84] for the simplicity and elegance of its implementation as well as for its outstanding convergence properties.

5 Solution using Split Bregman Algorithm

Directly solving the original problem (31) is difficult because of the compound nature of the regularization it involves. The split Bregman approach [50,51] allows reducing (31) to a simpler form through introduction on an auxiliary variable $u \in \mathfrak{V}$, which can be viewed as a noise-free version of the data field s . Particularly, using u (31) can be redefined as

$$\begin{aligned} \min_{c,u} \left\{ \frac{1}{2} \|u - s\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} + \mu \|u\|_{\mathfrak{V},\text{TV}} \right\} \\ \text{s.t. } \|\mathcal{A}\{c\} - u\|_{\mathfrak{V},2}^2 = 0. \end{aligned} \quad (36)$$

Then, starting from an arbitrary $b^0 \in \mathfrak{V}$, the Bregman algorithm [85] finds optimal c and u through the following iterations

$$(u^{t+1}, c^{t+1}) = \arg \min_{c, u} \left\{ \frac{1}{2} \|u - s\|_{\mathfrak{V}, 2}^2 + \lambda \|c\|_{\mathfrak{U}, 1} + \mu \|u\|_{\mathfrak{V}, \text{TV}} + \frac{\gamma}{2} \|u - \mathcal{A}\{c\} - b^t\|_{\mathfrak{V}, 2}^2 \right\} \quad (37)$$

$$b^{t+1} = b^t + (\mathcal{A}\{c^{t+1}\} - u^{t+1}),$$

for some $\gamma > 0^2$. The functional in (37) is supposed to be minimized over two variables, i.e. u and c . However, due to the way the ℓ_1 and TV components of this functional have been split, the minimization can now be performed efficiently by iteratively minimizing with respect to u and c separately. The resulting iteration steps are

$$\text{Step 1: } c^{t+1} = \arg \min_c \left\{ \frac{\gamma}{2} \|\mathcal{A}\{c\} - (u^t - b^t)\|_{\mathfrak{V}, 2}^2 + \lambda \|c\|_{\mathfrak{U}, 1} \right\} \quad (38)$$

$$\text{Step 2: } u^{t+1} = \arg \min_u \left\{ \frac{1}{2} \|u - s\|_{\mathfrak{V}, 2}^2 + \frac{\gamma}{2} \|u - (\mathcal{A}\{c^{t+1}\} + b^t)\|_{\mathfrak{V}, 2}^2 + \mu \|u\|_{\mathfrak{V}, \text{TV}} \right\}.$$

Note that the functional at Step 2 contains two quadratic terms which can be combined together to result in

$$\text{Step 2: } u^{t+1} = \arg \min_u \left\{ \frac{1 + \gamma}{2} \left\| u - \frac{s + \gamma (\mathcal{A}\{c^{t+1}\} + b^t)}{1 + \gamma} \right\|_{\mathfrak{V}, 2}^2 + \mu \|u\|_{\mathfrak{V}, \text{TV}} \right\}. \quad (39)$$

To cause a substantial reduction in the value of the cost functional in (37), Step 1 and Step 2 should be applied recursively for a predefined number of times before the Bregman parameter b^t is updated according to (37). It was argued in [51], however, that the extra precision gained through such a repetitive application of Step 1 and Step 2 is likely to be “wasted” when b^t is updated. Consequently, it was suggested in [51] to perform these steps only once per iteration cycle. It is interesting to note that, in this case, the split Bregman algorithm transforms into the *alternating directions method of multipliers* (ADMM) [52], whose convergence is guaranteed by the Eckstein-Bertsekas theorem [86] (see also Theorem 3.1 in [52]).

The final algorithm is summarized below. Lines 3-4 of Algorithm 1 correspond to Step 1 in (38), while lines 5-6 correspond to Step 2. An even more important fact to notice is that the optimization problem in line 4 is separable in the spatial coordinate \mathbf{r} . This optimization, therefore, can be performed at each voxel independently as discussed in Section 4.1. Moreover, the optimization problem in line 6 is separable in the diffusion coordinate k , and hence it amounts to applying TV-denoising to each of the K components of u independently as discussed in Section 4.2.

²Note that the algorithm is guaranteed to converge for any $\gamma > 0$. In this work we use $\gamma = 0.5$.

Algorithm 1 ADMM algorithm for sparse-TV reconstruction of HARDI signals

- 1: $b \leftarrow 0, u \leftarrow s$
 - 2: **while** “ c keeps changing” **do**
 - 3: $d \leftarrow u - b$
 - 4: $c \leftarrow \arg \min_c \left\{ \frac{1}{2} \|\mathcal{A}\{c\} - d\|_{\mathfrak{D},2}^2 + \frac{\lambda}{\gamma} \|c\|_{\mathfrak{U},1} \right\}$
 - 5: $d \leftarrow (1 + \gamma)^{-1} (s + \gamma (\mathcal{A}\{c\} + b))$
 - 6: $u \leftarrow \arg \min_u \left\{ \frac{1}{2} \|u - d\|_{\mathfrak{D},2}^2 + \frac{\mu}{1+\gamma} \|u\|_{\mathfrak{D},\text{TV}} \right\}$
 - 7: $b \leftarrow b + (\mathcal{A}\{c\} - u)$
 - 8: **end while**
-

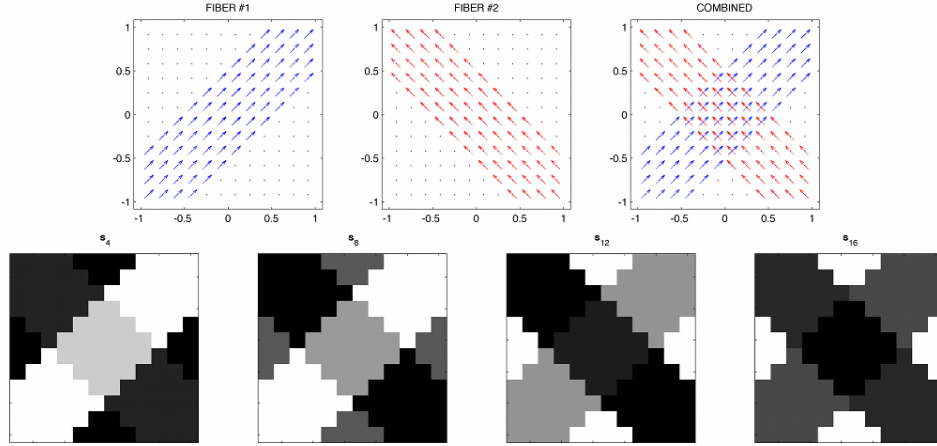


Figure 1: Phantom #1: (Upper row of subplots) The orientations of the individual diffusion flows and their combination; (Lower row of subplots) Examples of the resulting (noise-free) diffusion-encoding images corresponding to four different diffusion-encoding directions.

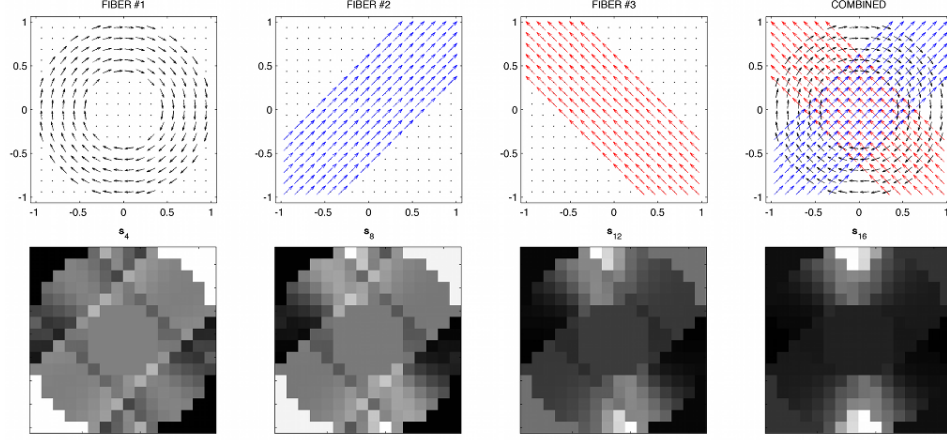


Figure 2: Phantom #2: (Upper row of subplots) The orientations of the individual diffusion flows and their combination; (Lower row of subplots) Examples of the resulting (noise-free) diffusion-encoding images corresponding to four different diffusion-encoding directions.

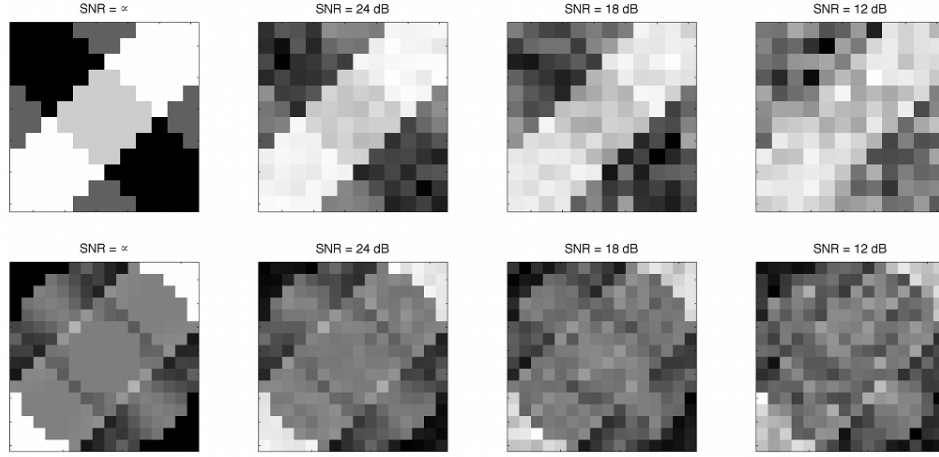


Figure 3: (Upper row of subplots) Diffusion-encoding images of Phantom #1 corresponding to $\mathbf{u} = [1, 1, 1]/\sqrt{3}$ and SNR = ∞ , 24, 18 and 12 dB; (Lower row of subplots) Diffusion-encoding images of Phantom #2 corresponding to the same \mathbf{u} and the same values of SNR.

6 Results

6.1 Technical details of the experimental study

Both the choice of diffusion-encoding directions $\{\mathbf{u}_k\}_{k=1}^K$ and of the orientations of spherical ridgelets require the use of a sampling scheme. In dMRI, one of the standard methods used to distribute a given number of spherical points over \mathbb{S}^2 in a quasi-uniform manner is by means of the electrostatic repulsion algorithm (also known as the Thomson problem). However, since the diffusion signals are symmetric (implying $s(\mathbf{u}|\mathbf{r}) = s(-\mathbf{u}|\mathbf{r})$), it is a unit hemisphere, not the entire \mathbb{S}^2 , which actually needs to be discretized. In view of the absence of a formulation of the Thomson problem for hemispherical domains, a common practice is to run the standard procedure for twice as many points as needed, followed by keeping only a half of the resulting configuration. However, as the retained points are not explicitly constrained to lie on a hemisphere, they may include nearly antipodal pairs which are likely to introduce undesirable dependencies between the diffusion measurements as well as between the basis functions. This limitation can be overcome through adapting a different sampling strategy. Particularly, in this paper, both the diffusion-encoding directions and the orientations of spherical ridgelets have been defined by using the method of generalized spiral points [87, 88], in which the sampling points are arranged along a spiral in such a way that the distance between the points along the spiral is approximately equal to the distance between its coils. This method is easily adaptable for sampling of the “northern” hemisphere (i.e. $\{\mathbf{u} \in \mathbb{S}^2 | \mathbf{u} \cdot [0, 0, 1]^T \geq 0\}$), providing a nearly uniform, unique and analytically computable coverage which is in no respect inferior to the one produced by solving the Thomson problem.

To assess the performance of the proposed algorithm under controllable conditions, experiments with simulated data sets have been performed. In this case, the HARDI signals were generated according to model (1) with different values of $M(\mathbf{r})$, $D_i(\mathbf{r})$, and $s_0(\mathbf{r}) = 1$, $\forall \mathbf{r}$. The resulting signals were contaminated by variable levels of Rician noise, giving rise to a set of different SNRs. In this work, we adapt the standard definition of the SNR as

$$\text{SNR} = 20 \log_{10} \left(\frac{\|s - \tilde{s}\|_{\mathfrak{H},2}}{\|s\|_{\mathfrak{H},2}} \right), \quad (40)$$

where s and \tilde{s} denote an original signal and its noise-contaminated version, respectively, and the norms are computed as defined by (18). It is worthwhile noting that the optimal values of regularization parameters λ and μ in (31) are normally a function of the noise level. In the present paper, however, no attempts have been extended to optimize these values for different SNRs. Instead, it was found that $\lambda = 0.03$ and $\mu = 0.05$ provided acceptable estimation results in all the simulation scenarios, and hence these values have been used throughout the whole study.

Following [37], the scaling parameter ρ in (6) was set to 0.5 and the resolution parameter J in (13) was set to be equal to 1, corresponding to a total of three resolution levels. The number of spherical ridgelet orientations were predefined

with $m_0 = 4$, resulting in $M_{-1} = 16$, $M_0 = 49$ and $M_1 = 169$ ridgelets spanning the resolution levels $j = -1$, $j = 0$ and $j = 1$, respectively. Thus, the total number of spherical ridgelets used in the reconstruction was equal to 234.

To quantitatively compare the reconstruction results produced by the proposed and references methods for different numbers of sampling directions K and various SNRs, three performance measures were used. The first of the three was the normalized mean-squared error (NMSE) defined as

$$\text{NMSE} = \frac{1}{N_x N_y N_z} \sum_{\mathbf{r} \in \Omega_d} \frac{\|s(\mathbf{r}) - \hat{s}(\mathbf{r})\|_2^2}{\|s(\mathbf{r})\|_2^2}, \quad (41)$$

with $s(\mathbf{r})$ being a reference HARDI signal corresponding to location \mathbf{r} and $\hat{s}(\mathbf{r})$ being its estimate. Depending on the nature of a specific experiment, the reference signal can be either a simulated signal discretized at 642 spherical points obtained by the 3rd order tessellation of the icosahedron *or* a signal reconstructed using a maximum possible number of diffusion-encoding orientations.

One of the most valuable outcomes of HARDI is in providing an access to computation of orientation distribution functions (ODFs) – the functions whose modes are likely to coincide with the direction of local diffusion flows [14]. Both the SH-based [28] and ridgelet-based [37] methods of reconstruction of HARDI signals come with analytical expressions which relate the HARDI signals to their corresponding ODFs. The latter can in turn be used to recover the directions of local diffusion flows (or, equivalently, the orientations of their related fibre tracts) using, e.g., the steepest ascent procedure detailed in [37]. Suppose \mathbf{u}_0 is the true direction of a diffusion flow and $\tilde{\mathbf{u}}$ is its estimate. Then, the angular orientation error δ can be defined (in degrees) as

$$\delta = \frac{180}{\pi} \arccos(\mathbf{u}_0 \cdot \tilde{\mathbf{u}}). \quad (42)$$

In this paper, as a performance measure, we use an *average* angular orientation error which is obtained by averaging the values of δ computed for all “fibres” within a specified Ω_d .

The last performance measure used in this work is the probability P_d of false fibre detection. To define P_d , let $M(\mathbf{r})$ be the true number of fibre tracts passing through voxel \mathbf{r} (as defined by model (1)). Also, let $\hat{M}(\mathbf{r})$ be an estimated number of fibres, which is equal to the number of modes (maxima) of the ODF recovered at position \mathbf{r} . Then, one can define

$$P_d = \left[\frac{1}{N_x N_y N_z} \sum_{\mathbf{r} \in \Omega_d} \frac{|M(\mathbf{r}) - \hat{M}(\mathbf{r})|}{M(\mathbf{r})} \right] \cdot 100\%. \quad (43)$$

In addition to the quantitative comparison, the reconstruction results will be evaluated through visual comparison as well. In this paper, we choose to visualize spherical functions by means of 3-D surface plots. Such a plot tends to project away from the origin of \mathbb{R}^3 in the directions along which a spherical

function is maximized, while passing near the origin in the directions where the function approaches zero.

Finally, our choice of reference methods was motivated by the scope of the main statements made in this paper. First, since we argue that the frame of spherical ridgelets is optimally suited for CS-based reconstruction of HARDI signals, its performance has to be compared with that of alternative representation systems. In particular, the basis of spherical harmonics up to the order 8 inclusive has been used for a different definition of the sensing matrix A in (14). (Note that, in the case of a real and symmetric analysis, this SH-basis consists of 45 functions.) Additionally, the representation system proposed in [46,47] has been exploited in the comparative study as well. This system is formed by applying a set of rotations to a Gaussian kernel of the form $d(\mathbf{u}) = \exp\{-b(\mathbf{u}^T D_0 \mathbf{u})\}$, with b defined as in (1) and D_0 equal to a (diagonal) diffusion tensor having a mean diffusivity of $766 \text{ mm}^2/\text{s}$ and a fractional anisotropy of 0.8. Following [47], the number of rotations (and hence the number of Gaussian basis functions) was set to be equal to 253. For the convenience of referencing, the CS-based reconstruction methods using the spherical ridgelets, the 8th-order spherical harmonics, and the rotated Gaussian kernels will be referred below to as the RDG, SH8 and GSS algorithms, respectively.

To assess the significance of the proposed spatial regularization, all the above algorithms have been applied with two different values of μ in (31), *viz.* $\mu = 0$ and $\mu = 0.05$. Note that, in the first of these cases, the spatial regularity is ignored, which leads to the sparse-only reconstruction discussed in Section 4.1. In the second case, on the other hand, the spatial regularity is taken into account and the reconstruction is performed by means of the split Bregman algorithm of Section 5.

6.2 In silico experiments

To assess the performance of the proposed and reference methods under controllable conditions, two simulated data sets were used. The first set (referred to below as Phantom #1) had a spatial dimension of 12×12 pixels, and consisted of two “fibres” crossing each other at the right angle as it is shown in the upper row of subplots of Fig. 1. In addition, each pixel in the set was assigned an extra diffusion flow in the direction perpendicular to the image plane. As a result, the number of diffusion components $M(\mathbf{r})$ in Phantom #1 varied between 1 and 3. Subsequently, model (1) was used to generate corresponding diffusion-encode images $\{s_k\}_{k=1}^K$ for a range of different values of K . Two different values of b , namely $b = 1000 \text{ s/mm}^2$ and $b = 3000 \text{ s/mm}^2$ were used for data generation. The diffusion tensors $D_i(\mathbf{r})$ in (1) were obtained by applying rotations to a tensor of the form $D_0 = \text{diag}([\alpha, \beta, \beta])$, where α and β were equal to $1700 \cdot 10^{-6}$ and $300 \cdot 10^{-6}$, respectively. Note that the mean diffusivity and fractional anisotropy of D_0 are equal to $766 \text{ mm}^2/\text{s}$ and 0.8, respectively. Thus, the same diffusion tensors were used for data synthesis and for the construction of basis functions in the GSS algorithm, thereby allowing the latter to perform under the best possible conditions.

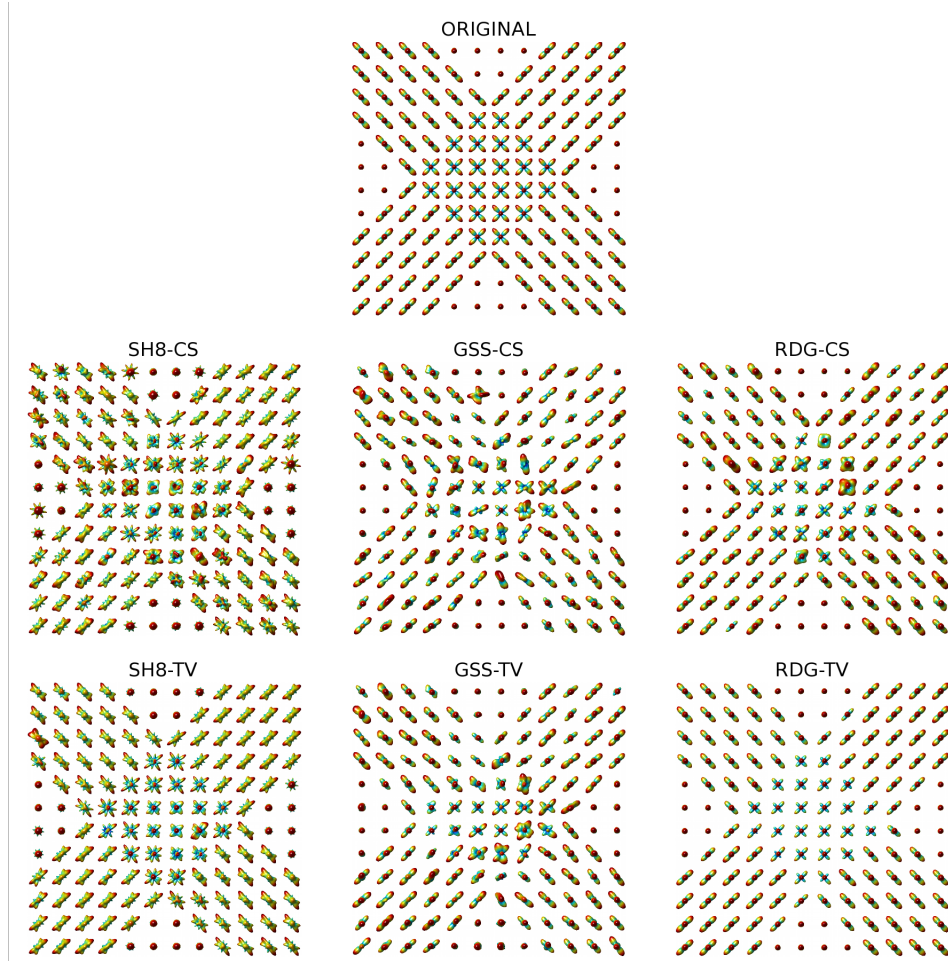


Figure 4: (Upper subplot) Original ODFs of Phantom #1; (Middle row of subplots) The ODFs recovered by the SH8-CS, GSS-CS, and RDG-CS algorithms, respectively; (Bottom row of subplots) The ODFs recovered by the SH8-TV, GSS-TV, and RDG-TV algorithms, respectively.

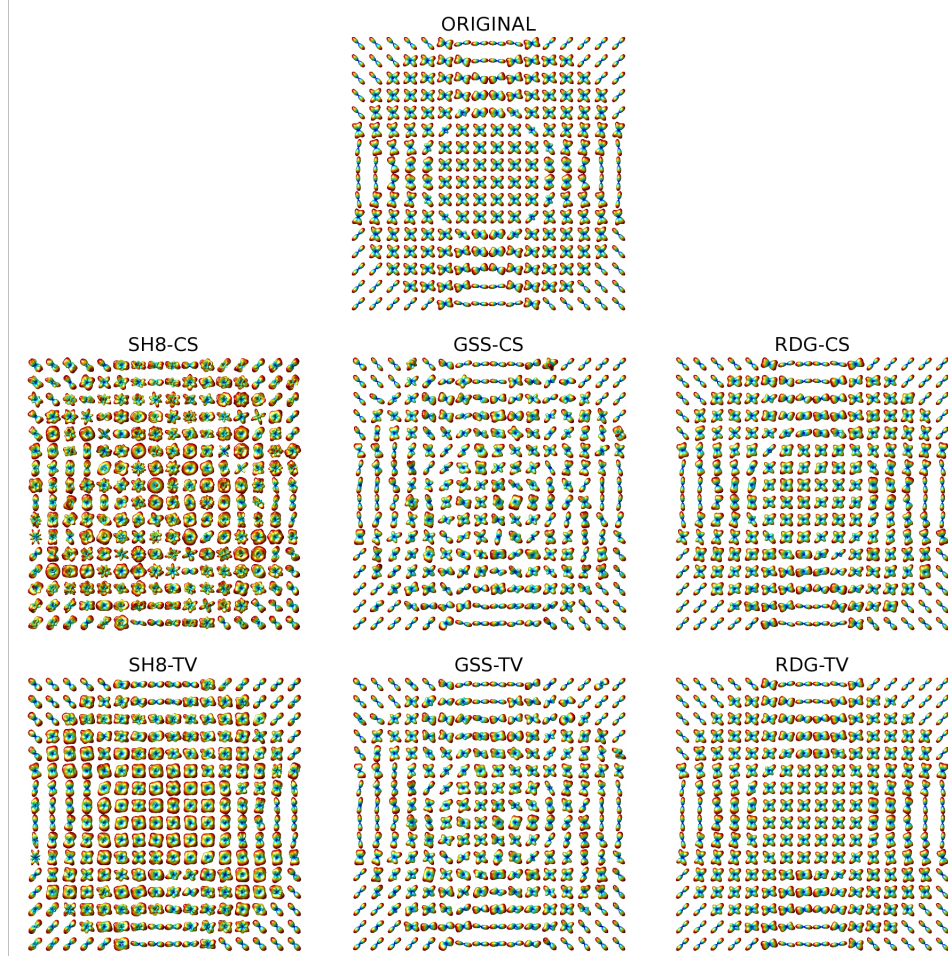


Figure 5: (Upper subplot) Original ODFs of Phantom #2; (Middle row of subplots) The ODFs recovered by the SH8-CS, GSS-CS, and RDG-CS algorithms, respectively; (Bottom row of subplots) The ODFs recovered by the SH8-TV, GSS-TV, and RDG-TV algorithms, respectively.

The lower row of subplots in Fig. 1 depict four examples of the diffusion-encoding images obtained for Phantom #1 before their contamination by Rician noise. One can see that the images are piecewise constant functions, which appears to be in a good agreement with the bounded-variation model suggested by (31). However, real images may be more complicated than that. Accordingly, to test the robustness of the proposed regularization scheme, a different *in silico* phantom was designed. Phantom #2 had a spatial dimension of 16×16 pixels and it was obtained through supplementing the configuration of Phantom #1 by an additional circular “fibre” as shown in the upper row of subplots in Fig. 2. The lower row of subplots of the figure show a subset of the resulting diffusion-encoded images, which can be seen to no longer exhibit a piecewise constant behaviour characteristic for Phantom #1.

The simulated diffusion-encoded images were contaminated by three different levels of Rician noise, giving rise to SNR of 24, 18 and 12 dB. Some typical examples of the resulting images are demonstrated in Fig. 3, where the upper row of subplots depict a noise-free version of one of the diffusion-encoded images pertaining to Phantom #1 along with its noise-contaminated counterparts. The lower row of subplots in Fig. 3 depict an analogous set of examples for Phantom #2. Observing the figure, one can see that the SNR values have been chosen so as to cover a range of possible noise scenarios, which could be characterized as moderate-to-severe contamination.

As it was mentioned earlier, in our *in silico* study we compared the performance of three different representation bases, i.e. spherical harmonics (SH8), Gaussian kernels (GSS) and spherical ridgelets (RDG). All the resulting algorithms have been further subdivided into two different types, depending on whether or not the spatial regularization was engaged. Thus, in the absence of the spatial regularization (corresponding to $\mu = 0$), the reconstruction has been performed on a voxel-by-voxel basis, as detailed in Section 4.1. For the convenience of referencing, the corresponding algorithms will be referred to below as SH8-CS, GSS-CS, and RDG-CS. In the case of $\mu > 0$, the estimation has been carried out using the split Bregman method of Section 5. The corresponding algorithms will be referred below as SH8-TV, GSS-TV, and RDG-TV.

The upper subplot of Fig. 4 shows the original field of ODFs of Phantom #1 (corresponding to $b = 3000$ s/mm²), which have been computed based on Tuch’s approximation [14] (i.e. by applying the Funk-Radon transform to the diffusion signals). At the same time, the middle row of subplots of Fig. 4 show the ODFs recovered by (from left to right) SH8-CS, GSS-CS and RDG-CS with $K = 16$ and SNR = 18 dB. One can see that the inability of the SH basis to sparsely represent HARDI signals results in a poor performance of SH8-CS. A better result is obtained with GSS-CS, which uses a basis of rotated Gaussian kernels, and therefore has a potential to represent the HARDI signals in a sparse manner. Unfortunately, the excessive correlation between the Gaussian basis functions adversely affects the ability of this method to withstand the effect of noise. Consequently, the reconstruction obtained using GSS-TV suffers from sizeable errors. The RDG-CS method, on the other hand, provides an estimation result of a much higher quality, albeit some inaccuracies are still noticeable in the

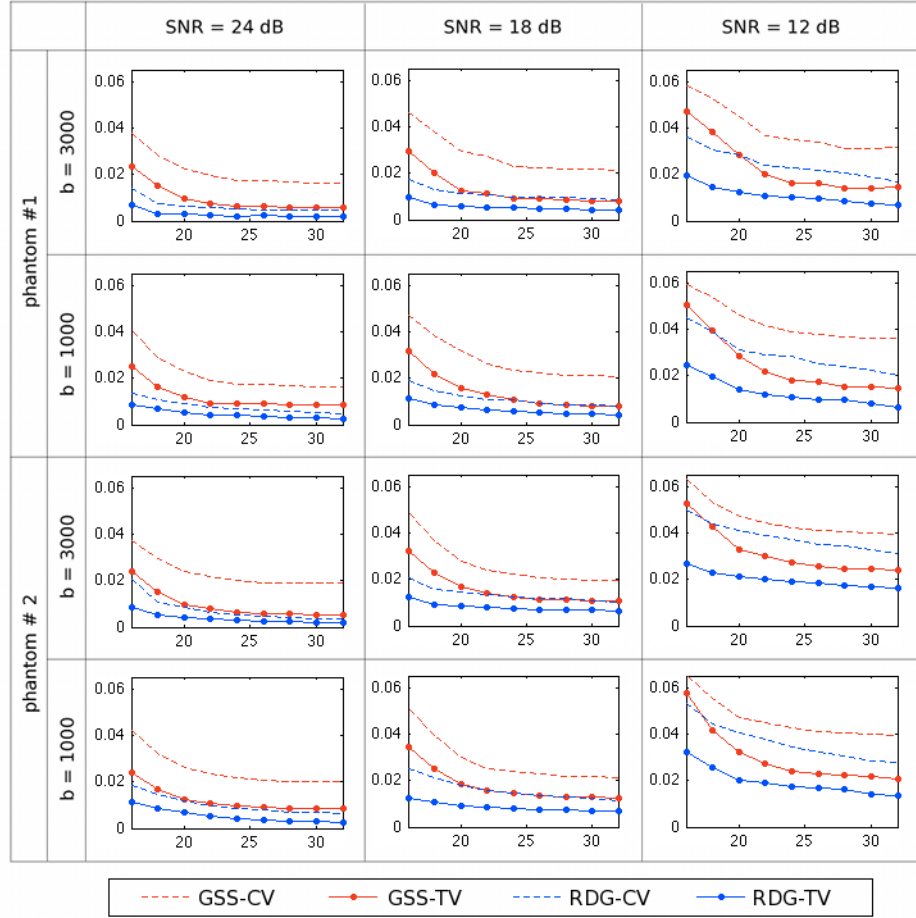


Figure 6: NMSE obtained using the compared methods for different phantoms, SNRs and b -values.

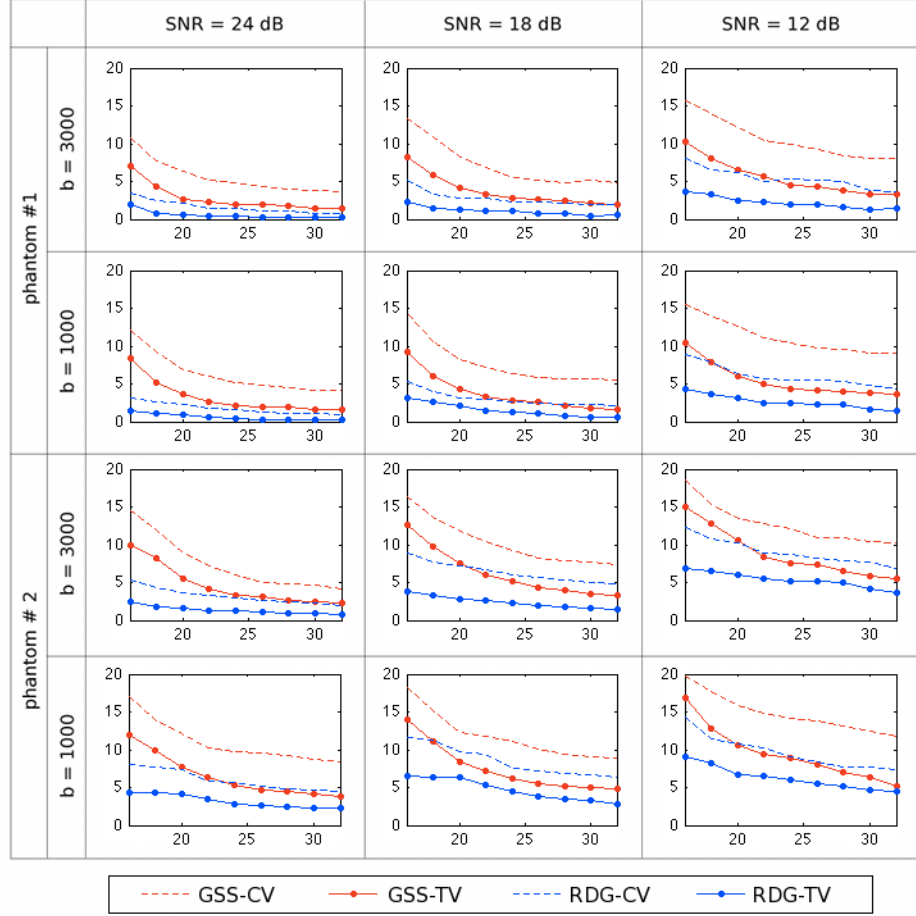


Figure 7: Average angular error δ obtained using the compared methods for different phantoms, SNRs and b -values.

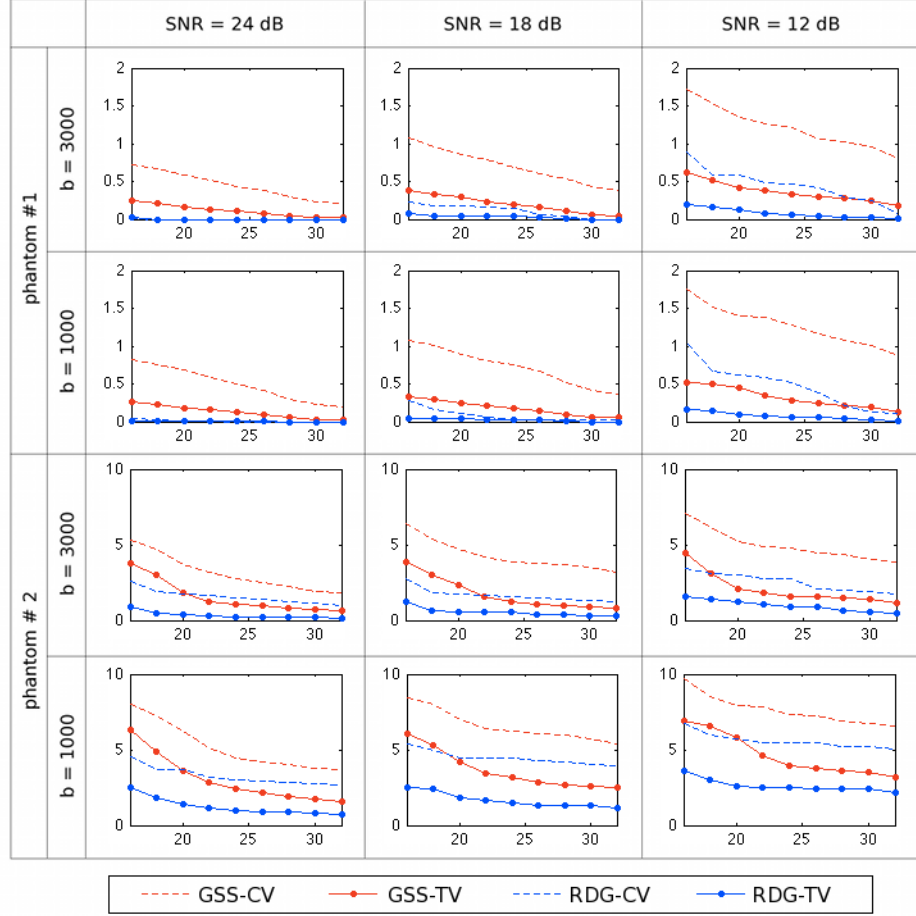


Figure 8: The rate of false fibre detection P_d obtained using the compared methods for different phantoms, SNRs and b -values.

central part of the phantom. The reconstruction accuracy improves dramatically when the spatial regularization is “switched on”, as it is demonstrated by the bottom row of subplots in Fig. 4. Specifically, while SH8-TV is still unable to provide a valuable reconstruction, the estimates obtained using GSS-TV and RDG-TV represent correctly the “flow structure” of Phantom #1. Moreover, among the latter two methods, RDG-TV is clearly the best performer, resulting in a close-to-ideal recovery of the original ODFs. The superiority of RDG-TV over the alternative methods is further evident in the results presented by Fig. 5, which depicts the reconstructions obtained for Phantom #2 (with the same values of b , K and SNR as above).

In general, the reconstruction results obtained using SH8-CS and SH8-TV have been observed to be of a lower quality in comparison to the other methods under consideration. For this reason, in what follows, only the GSS and RDG methods are compared. Thus, Fig. 6 contrasts the performances of GSS-CS, GSS-TV, RDG-CS and RDG-TV in terms of the NMSE criterion. One can see that the best performance here is attained by the RDG-TV algorithm, which results in the smallest values of NMSE for both phantoms and for all the tested values of b , SNR and K . It is also interesting to note that the incorporation of spatial regularization allows GSS-TV to outperform RDG-CS, with the effect of the regularization becoming more pronounced at lower SNRs. On the whole, all the NMSE curves demonstrate an expected behaviour, with the error values increasing proportionally with a decrease in SNR, while going down with an increase in the number of diffusion-encoding gradients K . However, as opposed to the others, the NMSE curves obtained with RDG-TV are characterized by a relatively low rate of convergence, which indicates a reduced sensitivity of RDG-TV to the value of K .

The above algorithms have been also compared in terms of the angular error (42). The results of this comparison are summarized in Fig. 7, which again indicates that the most accurate reconstruction is obtained using the RDG-TV method. In general, the angular error tends to grow as SNR decreases and to converge to a minimum as K increases. As opposed to the case of NMSE, however, there is an additional dependency of the angular error on the type of a phantom in use as well as on the b -value. In particular, the errors obtained for Phantom #2 are (on average) greater than those obtained for Phantom #1. This discrepancy is rooted in the fact that Phantom #2 has a more complex “fibre structure” as compared to Phantom #1. In particular, while the “fibers” of Phantom #1 are designed to cross each other at the right angle, the “fibres” of Phantom #2 are allowed to decussate at much smaller angles, which makes them much harder to resolve. Moreover, this effect becomes more noticeable with a decrease in the b -value, which reduces the resolution of q-ball imaging. Finally, we notice that, on average, GSS-TV performs better than RDG-CS (though still worse than RDG-TV), which justifies the value of spatial regularization.

The comparison in terms of the rate of false fibre detection P_d (43) was last in the line of our *in silico* performance tests; its results are shown in Fig. 8. One can see that, in the case of Phantom #1, RDG-TV yields a virtually zero false detection rate for both values of b , whereas the other methods result in

considerably higher values of P_d (mainly due to the detection of spurious local maxima in the estimated ODFs). The situation is different for Phantom #2, where all the compared methods yield sizeable errors (especially for $b = 1000$ s/mm²). However, in comparative terms, the most accurate reconstruction is still obtained by means of the proposed RDG-TV algorithm.

6.3 In vivo results

As the next validation step, experiments with real HARDI data were carried out. The proposed algorithm was tested on human brain scans acquired on a 3-Tesla GE system using an echo planar imaging (EPI) diffusion-weighted image sequence. A double echo option was used to suppress eddy-current related distortions. To improve the spatial resolution of EPI, an eight channel coil was used to perform parallel imaging by means of the ASSET technique with a speed-up factor of 2. The data were acquired using 51 gradient directions (quasi-uniformly distributed over the northern hemisphere) with $b = 1000$ s/mm². In addition, eight baseline (b_0) scans were acquired, averaged and used for normalization. The following scanning parameters were used: TR = 17000 ms, TE = 78 ms, FOV = 24 cm, 144×144 encoding steps, and 1.7 mm slice thickness. All scans had 85 axial slices parallel to the AC-PC line covering the whole brain.

The main question addressed through the *in vivo* experiments has been whether or not it is possible to supersede the spatial regularization by pre-filtering of HARDI signals. To this end, the RDG-CS algorithm was applied first to the HARDI data containing the full set of $K = 51$ diffusion gradients. (Note that such *dense* reconstruction is analogous to the one reported in [37], where the latter is shown to outperform the SH-based estimation [28].) The resulting ODFs have been used as a fiducial against which different reconstruction results were compared.

As the next step, three different subsets of 16, 24 and 32 spherical points were composed out of the original set of 51 diffusion gradients. Within each of these subsets, their corresponding points were chosen so as to result in a quasi-uniform coverage of the northern hemisphere. Accordingly, the HARDI data were rearranged into three data sets of size $144 \times 144 \times 85 \times 16$, $144 \times 144 \times 85 \times 24$ and $144 \times 144 \times 85 \times 32$ to emulate compressed sensing data acquisition. The above sets were used to assess the performance of different reconstruction methods. Unfortunately, we have not succeeded to find conditions under which the SH8 and GSS algorithms would provide stable reconstruction results (either with or without pre-filtering). For this reason, only the RDG-CS and RDG-TV algorithms are compared below.

The upper row of subplots in Fig. 9 show the generalized anisotropy (GA) [14] image of a coronal cross-section of the brain along with the reference field of ODFs corresponding to the region indicated by the yellow rectangular. Anatomically, this region is expected to contain the fibre bundles of corona radiata as well as those of superior longitudinal and arcuate fasciculi. The middle row of subplots in the same figure depict the ODFs reconstructed by RDG-CS using

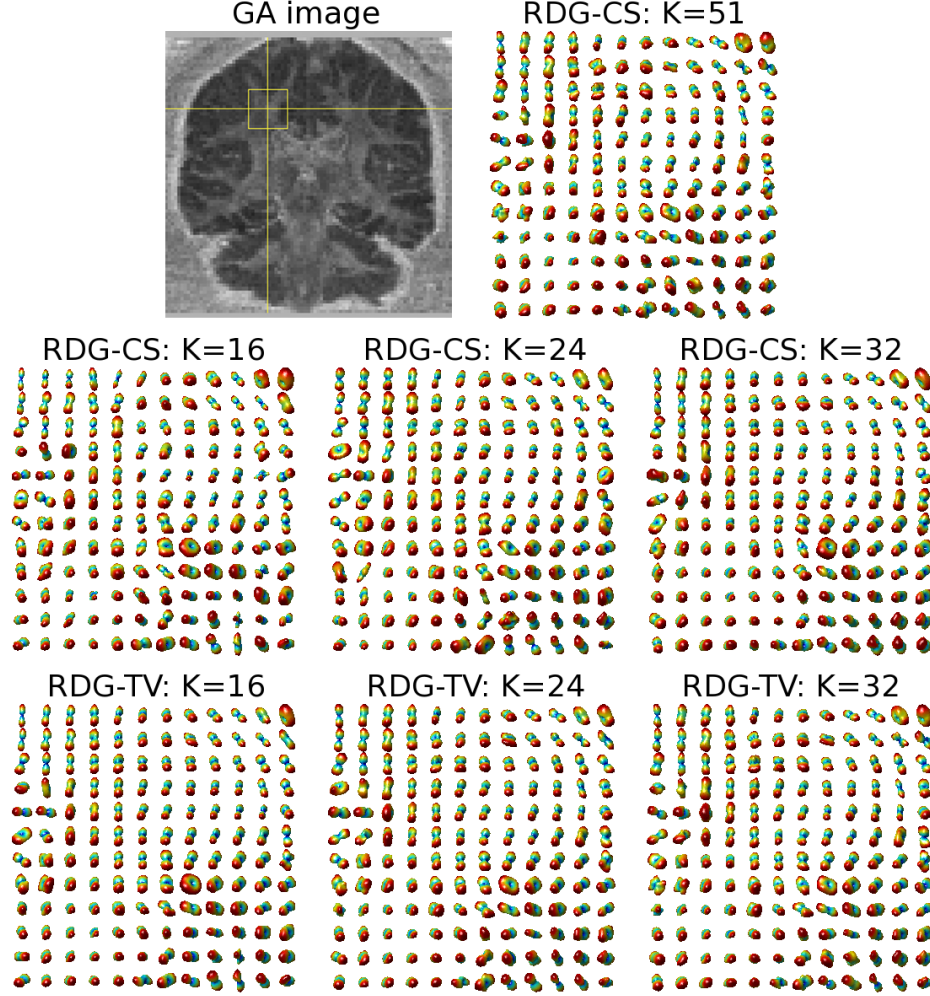


Figure 9: (*Upper row of subplots*) A coronal GA image and the ODF field of the indicated region recovered by RDG-CS with $K = 51$; (*Middle row of subplots*) Estimated ODF fields obtained using RDG-CS with $K = 16$, $K = 24$ and $K = 32$; (*Bottom row of subplots*) Estimated ODF fields obtained using RDG-TV with $K = 16$, $K = 24$ and $K = 32$.

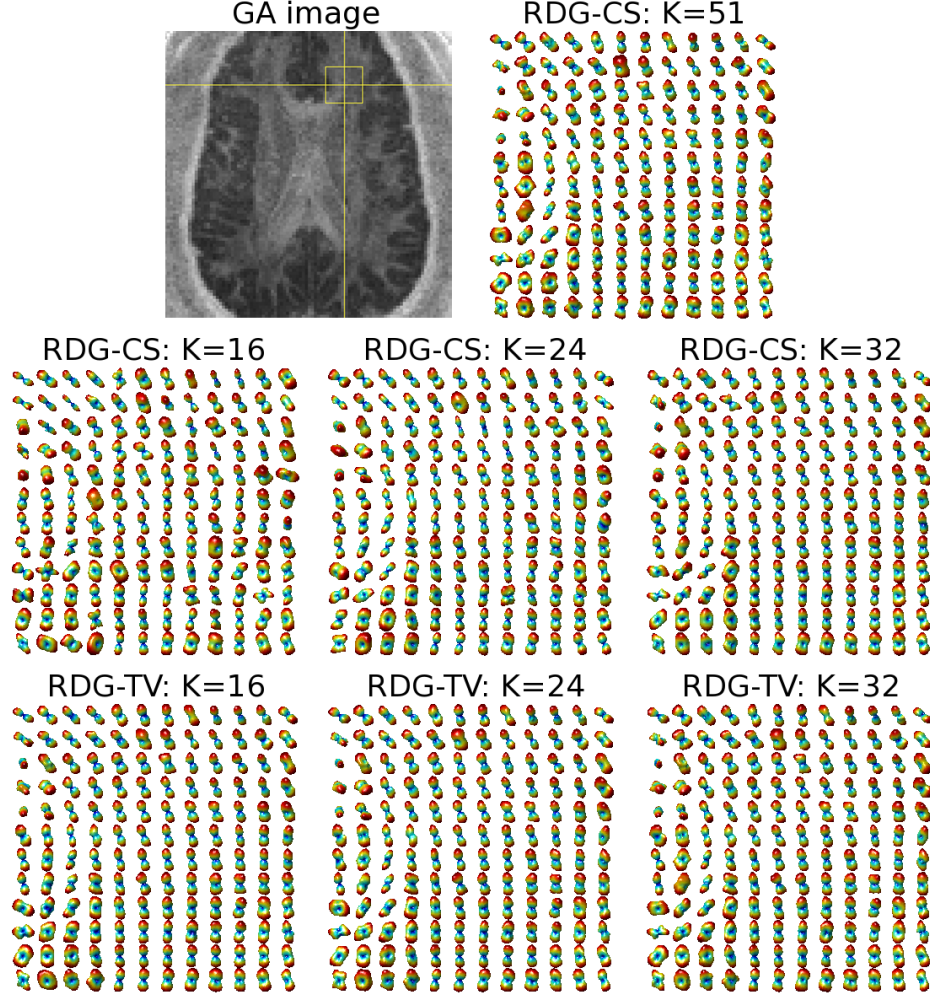


Figure 10: (*Upper row of subplots*) An axial GA image and the ODF field of the indicated region recovered by RDG-CS with $K = 51$; (*Middle row of subplots*) Estimated ODF fields obtained using RDG-CS with $K = 16$, $K = 24$ and $K = 32$; (*Bottom row of subplots*) Estimated ODF fields obtained using RDG-TV with $K = 16$, $K = 24$ and $K = 32$.

Table 1: NMSE computed between the dense and CS-based reconstructions obtained with RDG-CS and RDG-TV

| | Pertaining to Fig. 9 | | | Pertaining to Fig. 10 | | |
|--------|----------------------|----------|----------|-----------------------|----------|----------|
| | $K = 16$ | $K = 24$ | $K = 32$ | $K = 16$ | $K = 24$ | $K = 32$ |
| RDG-CS | 0.097 | 0.064 | 0.043 | 0.091 | 0.053 | 0.037 |
| RDG-TV | 0.022 | 0.011 | 0.003 | 0.018 | 0.009 | 0.002 |

$K = 16, 24$ and 32 diffusion gradients. One can see that the quality of reconstruction progressively improves as K increases. It is important to note that, before applying the RDG-CS algorithm, the diffusion-encoded images had been pre-processed by a TV filter to reduce the effect of measurement noises on the estimation result. However, this pre-processing appears to be not nearly as effective as the spatial regularization of the RDG-TV algorithm, whose reconstruction results are shown in the bottom row of subplots in Fig. 9. The above conclusion is further supported by an additional example of Fig. 10, which shows the reconstructions pertaining to the indicated area within an axial cross-section of the brain. (The relevant fibre bundles here are those of cingulum and corpus callosum). As in the previous example, one can see that the most accurate reconstruction is attained by means of the proposed RDG-TV method. The superiority of RDG-TV is also confirmed by the quantitative figures of Table 1, which summarizes the NSME obtained by the compared algorithms for different values of K .

7 Discussion and Conclusions

When considered as a whole, the HARDI signals which pertain to a given volume of interest can be described as multi-valued (or, more generally, measure-valued) functions from a subset of \mathbb{R}^3 to the space of square-integrable spherical functions $\mathbb{L}_2(\mathbb{S}^2)$. Such functions can be thought of as if they had two “modes of variation” - one in the spatial and another in the diffusion domain. Although applying various inverse problems (a particular instance of which is addressed by the theory of CS) along the spatial and diffusion coordinates independently is not new to the community of medical imaging scientists, formulating a CS reconstruction problem in *both domains simultaneously* has not been proposed before. Accordingly, the present paper introduced the RDG-TV algorithm which exploits the above idea and can be used for reliable reconstruction of HARDI signals from as few as $K = 16$ diffusion-encoded scans (as compared to 60-80 scans required by existing reconstruction tools). The algorithm exploits the fact that HARDI signals can be sparsely represented by spherical ridgelets in the diffusion domain, while their associated diffusion-encoded images have bounded variation in the spatial domain. Moreover, it has been shown experimentally that either using different representation bases or excluding the spatial regularization would result in considerably less accurate reconstruction results.

At the practical level, the reconstruction is implemented based on the split

Bregman approach (with 20 being the maximum number of Bregman iterations used in this study). The resulting algorithm alternates between two estimation stages (38): first, a sequence of basis pursuit de-noising problems are solved independently on a voxel-by-voxel basis, followed by applying a TV filter to a total of K discrete images. Such computations are straightforward to accelerate using a multicore processing, which is another advantage of the proposed reconstruction method.

We believe that the algorithm presented in this paper can be improved in a number of ways. First, the square metric used to assess the model fidelity could be replaced by a different metric, which would be more specific to the nature of Rician noise. Second, the fact that diffusion signals are positive-valued could be explicitly incorporated into the reconstruction process in the form of additional constraints. Lastly, the bounded variation model could be substituted by an alternative model, which could (possibly) provide a better account for the spatial regularity of HARDI signals. Exploring the above options constitutes essential part of our ongoing research.

Finally, as the experimental study reported in this paper was comparative in its nature, it was not really important what method to use for approximation of ODFs. Specifically, the present results have been obtained using Tuch’s approximation [14]. However, more accurate computation of ODFs is possible based on the solid angle formulation as detailed in [89, 90]³. It should also be noted that the technique proposed in [89] can be applied to multi-shell HARDI data (i.e. the data acquired for a range of b -values). Until recently, collecting such data has been deemed impractical due to extremely long acquisition time required. We believe, however, that the proposed method for CS-based reconstruction of HARDI data has a potential to help multi-shell HARDI develop into a clinically relevant tool of diagnostic imaging.

References

- [1] T. M. Jessell and E. R. Kandel, “Synaptic transmission: a bidirectional and self-modifiable form of cell-cell communication,” *Cell*, vol. 72 Suppl, pp. 1–30, Jan 1993.
- [2] H. Johansen-Berg and T. E. J. Behrens, *Diffusion MRI: From quantitative measurements to in-vivo neuroanatomy*, 1st ed. Academic Press, 2009.
- [3] D. L. Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet, “MR imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurological disorders,” *Radiology*, vol. 161, pp. 401–407, 1986.
- [4] P. J. Basser, J. Mattiello, and D. LeBihan, “MR diffusion tensor spectroscopy and imaging,” *Biophys. J.*, vol. 66, no. 1, pp. 259–267, 1994.

³We found, however, that the methods in [89, 90] are more sensitive to noise than the one in [14], and this is why we used the latter to avoid an inconsistency in interpreting the estimation results.

- [5] —, “Estimation of the effective self-diffusion tensor from the NMR spin echo,” *J. Magn. Reson. Imag. B*, vol. 103, no. 3, pp. 247–254, 1994.
- [6] C. Pierpaoli, P. Jezzard, P. Basser, A. Barnett, and G. D. Chiro, “Diffusion tensor MR imaging of the human brain,” *Radiology*, vol. 201, no. 3, pp. 637–648, 1996.
- [7] D. L. Bihan, J.-F. Mangin, C. Poupon, C. Clark, S. Pappata, N. Molko, and H. Chabriet, “Diffusion tensor imaging: Concepts and applications,” *Journal of Magnetic Resonance Imaging*, vol. 13, pp. 534–546, 2001.
- [8] S. Mori, K. Frederiksen, P. C. M. V. Zijl, B. Stieltjes, M. A. Kraut, M. Solaiyappan, and M. G. Pomper, “Brain white matter anatomy of tumor patients evaluated with diffusion tensor imaging,” *Annals of Neurology*, vol. 51, no. 3, pp. 377–380, 2002.
- [9] G. K. Rohde, A. S. Barnett, P. J. Basser, S. Marenco, and C. Pierpaoli, “Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI,” *Magnetic Resonance in Medicine*, vol. 51, pp. 103–114, 2004.
- [10] C. A. Clark, T. R. Barrick, M. M. Murphy, and B. A. Bell, “White matter fiber tracking in patients with space-occupying lesions of the brain: A new technique for neurosurgical planning?” *NeuroImage*, vol. 20, no. 3, pp. 1601–1608, 2003.
- [11] H. Akai, H. Mori, S. Aoki, Y. Masutani, N. Kawahara, J. Shibahara, and K. Ohtomo, “Diffusion tensor tractography of gliomatosis cerebri: Fiber tracking through the tumor,” *Journal of Computer Assisted Tomography*, vol. 29, no. 1, pp. 127–129, 2005.
- [12] C. S. Yu, K. C. Li, Y. Xuan, X. M. Ji, and W. Qin, “Diffusion tensor tractography in patients with cerebral tumors: A helpful technique for neurosurgical planning and postoperative assessment,” *European Journal of Radiology*, vol. 56, no. 2, pp. 197–204, 2005.
- [13] D. C. Alexander, “Multiple-fiber reconstruction algorithms for diffusion MRI,” *Annals of the New York Academy of Science*, vol. 1064, pp. 113–133, 2005.
- [14] D. S. Tuch, “Q-ball imaging,” *Magnetic Resonance in Medicine*, vol. 52, pp. 1358–1372, 2004.
- [15] S. Mori, *Introduction to diffusion tensor imaging*. Elsevier, 2007.
- [16] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi, “In vivo fiber tractography using DT-MRI data,” *Magnetic Resonance in Medicine*, vol. 44, pp. 625–632, 2000.

- [17] A. L. Alexander, J. E. Lee, M. Lazar, and A. S. Field, “Diffusion tensor imaging of the brain,” *Neurotherapeutics*, vol. 4, pp. 316–329, 2007.
- [18] D. K. Jones, “The effect of gradient sampling schemes on measures derived from Diffusion Tensor MRI: A Monte Carlo study,” *Magnetic Resonance in Medicine*, vol. 51, pp. 807–815, 2004.
- [19] M. R. Wiegell, H. Larsson, and V. J. Wedeen, “Fiber crossing in human brain depicted with diffusion tensor MR imaging,” *Radiology*, vol. 217, pp. 897–903, 2000.
- [20] L. Frank, “Anisotropy in high angular resolution diffusion-tensor MRI,” *Magnetic Resonance in Medicine*, vol. 45, pp. 935–939, 2001.
- [21] D. C. Alexander, G. J. Barker, , and S. R. Arridge, “Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data,” *Journal of Magnetic Resonance Imaging*, vol. 48, no. 2, pp. 331–340, 2002.
- [22] L. R. Frank, “Characterization of anisotropy in high angular resolution diffusion-weighted MRI,” *Magnetic Resonance in Medicine*, vol. 47, pp. 1083–1099, 2002.
- [23] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen, “High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity,” *Magnetic Resonance in Medicine*, vol. 48, pp. 577–582, 2002.
- [24] D. S. Tuch, T. G. Reese, M. R. Wiegell, and V. J. Wedeen, “Diffusion MRI of complex neural architecture,” *Neuron*, vol. 40, pp. 885–895, 2003.
- [25] A. W. Anderson, “Measurement of fiber orientation distributions using high angular resolution diffusion imaging,” *Magnetic Resonance in Medicine*, vol. 54, pp. 1194–1206, 2005.
- [26] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, “Apparent diffusion coefficients from high angular resolution diffusion images: Estimation and applications,” *Magnetic Resonance in Medicine*, vol. 56, no. 2, pp. 395–410, 2006.
- [27] C. P. Hess, P. Mukherjee, E. T. Han, D. Xu, and D. R. Vigneron, “Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis,” *Magnetic Resonance in Medicine*, vol. 56, pp. 104–117, 2006.
- [28] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, “Regularized, fast, and robust analytical Q-ball imaging,” *Magnetic Resonance in Medicine*, vol. 58, pp. 497–510, 2007.

- [29] B. W. Kreher, J. F. Schneider, I. Mader, E. Martin, J. Hennig, and K. A. Ilyasov, “Multi-tensor approach for analysis and tracking of complex fiber configurations,” *Magnetic Resonance in Medicine*, vol. 54, pp. 1216–1225, 2005.
- [30] A. Banerjee, I. Dhillon, J. Gosh, and S. Sra, “Clustering on the unit hypersphere using von Mises-Fisher distributions,” *J. Mach. Learn. Research*, vol. 6, pp. 1345–1382, 2005.
- [31] T. McGraw, B. Vemuri, B. Yezierski, and T. Mareci, “Von Mises-Fisher mixture model of the diffusion ODF,” in *Proceedings of ISBI*, Arlington, VA, 2006.
- [32] T. Schultz and H. Seidel, “Estimating crossing fibers: A tensor decomposition approach,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1635–1642, 2008.
- [33] E. Kaden, A. Anwender, and T. R. Knosche, “Variational inference of the fiber orientation density using diffusion mr imaging,” *NeuroImage*, vol. 42, pp. 1366–1380, 2008.
- [34] Y. Rathi, O. Michailovich, M. E. Shenton, and S. Bouix, “Directional functions for orientation distribution estimation,” *Medical Image Analysis*, vol. 13, no. 3, pp. 432–444, 2009.
- [35] H. Groemer, *Geometric applications of Fourier series and spherical harmonics*. Cambridge University Press, 1996.
- [36] O. Michailovich and Y. Rathi, “Approximation of orientation distribution functions by spherical ridgelets,” in *Proceedings of ISBI*, Paris, 2008.
- [37] —, “On approximation of orientation distributions by means of spherical ridgelets,” *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 461–477, 2010.
- [38] E. Candes and T. Tao, “Near optimal signal recovery from random projections: Universal encoding strategies,” *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [39] E. Candes, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1221, 2006.
- [40] —, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [41] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

- [42] D. Donoho and Y. Tsaig, “Extensions of compressed sensing,” *Signal Processing*, vol. 86, no. 3, pp. 533–548, 2006.
- [43] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4036–4048, 2006.
- [44] R. Baraniuk, M. Davenport, R. DeVore, , and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [45] O. Michailovich and Y. Rathi, “Fast and accurate reconstruction of HARDI data using compressed sensing,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Beijing, China, September 2010.
- [46] B. A. Landman, H. Wan, J. A. Bogovic, P.-L. Bazin, and J. L. Prince, “Resolution of crossing fibers with constrained compressed sensing using traditional Diffusion Tensor MRI,” in *Proceedings of SPIE*, ser. Medical Imaging, San Diego, CA, February 2010.
- [47] B. A. Landman, H. Wan, J. A. Bogovic, P. C. M. V. Zijl, P.-L. Bazin, and J. L. Prince, “In the pursuit of intra-voxel fiber orientations: Comparison of compressed sensing DTI and Q-ball MRI,” in *Proceedings of ISMRM*, Stockholm, Sweden, 2010.
- [48] M. I. Menzel, K. Khare, K. F. King, X. Tao, C. J. Hardy, and L. Marinelli, “Accelerated diffusion spectrum imaging in the human brain using compressed sensing,” in *Proceedings of ISMRM*, Stockholm, Sweden, 2010.
- [49] N. Lee and M. Singh, “Compressed sensing based Diffusion Spectrum Imaging,” in *Proceedings of ISMRM*, Stockholm, Sweden, 2010.
- [50] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, “Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing,” *SIAM J. Imaging Sciences*, vol. 1, no. 1, pp. 143–168, 2008.
- [51] T. Goldstein and S. Osher, “The split Bregman algorithm for L1 regularized problems,” UCLA, Tech. Rep. CAM Report 08-29, 2008.
- [52] E. Esser, “Applications of Lagrangian-based alternating direction methods and connections to split Bregman,” UCLA, Tech. Rep. CAM Report 09-31, 2008.
- [53] J. Malcolm, M. Shenton, and Y. Rathi, “Neural tractography using an unscented Kalman filter,” in *Proceedings of IPMI*, 2009, pp. 126–138.
- [54] —, “Filtered tractography: State estimation in a constrained subspace,” in *Proceedings of MICCAI (Diffusion Modeling and Fiber Cup)*, 2009, pp. 122–133.

- [55] K. M. Jansons and D. C. Alexander, “Persistent angular structure: New insights from diffusion magnetic resonance imaging data,” *Inverse Problems*, vol. 19, pp. 1031–1046, 2003.
- [56] J.-D. Tournier, F. Calamante, D. G. Gadian, and A. Connelly, “Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution,” *NeuroImage*, vol. 23, pp. 1176–1185, 2004.
- [57] A. Ramirez-Manzanares, M. Rivera, B. C. Vemuri, P. Carney, and T. Mareci, “Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry,” *IEEE Transactions on Medical Imaging*, vol. 26, no. 8, pp. 1091–1102, August 2007.
- [58] K. E. Sakaie and M. J. Lowe, “An objective method for regularization of fiber orientation distributions derived from diffusion-weighted MRI,” *NeuroImage*, vol. 34, pp. 169–176, 2007.
- [59] F. Huang, J. Akao, S. Vijayakumar, G. R. Duensing, and M. Limkeman, “k-t GRAPPA: A k-space implementation for dynamic MRI with high reduction factor,” *Magnetic Resonance in Medicine*, vol. 54, no. 5, pp. 1172–1184, 2005.
- [60] H. Jung, J. C. Ye, and E. Y. Kim, “Improved k-t BLASK and k-t SENSE using FOCUSS,” *Physics in Medicine and Biology*, vol. 52, pp. 3201–3226, 2007.
- [61] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, “k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI,” *Magnetic Resonance in Medicine*, vol. 61, pp. 103–116, 2009.
- [62] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [63] Y.-C. Kim, S. S. Narayanan, and K. S. Nayak, “Accelerated three-dimensional upper airway MRI using compressed sensing,” *Magnetic Resonance in Medicine*, vol. 61, pp. 1434–1440, 2009.
- [64] R. Chartrand, “Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data,” in *Proceedings of ISBI*, Boston, MA, 2009.
- [65] J. Trzasko, A. Manduca, and E. Borisch, “Highly undersampled magnetic resonance image reconstruction via homotopic L0-minimization,” *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2009.
- [66] D. Liang, B. Liu, J. Wang, and L. Ying, “Accelerating SENSE using compressed sensing,” *Magnetic Resonance in Medicine*, vol. 62, pp. 1574–1584, 2009.

- [67] I. Daubechies, *Ten lectures on wavelets*. SIAM, 1995.
- [68] S. Mallat, *A wavelet tour of signal processing*, 2nd ed. Academic Press, 1999.
- [69] W. Freeden, M. Schreiner, and R. Franke, “A survey on spherical spline approximation,” *Surv. Math. Ind.*, vol. 7, pp. 29–85, 1997.
- [70] W. Freeden and F. Schreiner, “Orthogonal and non-orthogonal multiresolution analysis, scale discrete and exact fully discrete wavelet transform on the sphere,” *Constructive Approximation*, vol. 14, pp. 493–515, 1998.
- [71] P. Blomberg and T. Chan, “Color tv: Total variation methods for restoration of vector-valued images,” *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 304–309, March 1998.
- [72] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [73] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- [74] I. Daubechies, M. Defrise, and C. DeMol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *arXiv:math/0307152v2*, 2003.
- [75] M. A. T. Figueiredo and R. D. Nowak, “An EM algorithm for wavelet-based image restoration,” *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.
- [76] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky, “A wide-angle view at iterated shrinkage algorithms,” in *Proceedings of SPIE (Wavelet XII)*, San-Diego CA, USA, 2007.
- [77] J. M. Bioucas-Dias and M. A. T. Figueiredo, “A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration,” *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.
- [78] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [79] C. Vogel and M. Oman, “Iterative methods for total variation de-noising,” *SIAM J. Sci. Comput.*, vol. 17, no. 1–4, pp. 227–238, Jan. 1996.
- [80] J.-F. Aujol, “Some first-order algorithms for total variation based image restoration,” *J. Math. Imaging Vis.*, vol. 34, pp. 307–327, 2009.

- [81] Y. Wang, W. Yin, and Y. Zhang, “A fast algorithm for image deblurring with total variation regularization,” Rice University, Tech. Rep. TR07-10, 2010.
- [82] O. Michailovich, “An iterative shrinkage approach to total-variation image restoration,” *IEEE Transactions on Image Processing*, 2010.
- [83] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, “An iterative regularization method for total variation-based image restoration,” *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 460–489, 2005.
- [84] A. Chambolle and P.-L. Lions, “Image recovery via total variation minimization and related problems,” *Numer. Math.*, vol. 76, pp. 167–188, 1997.
- [85] L. M. Bregman, “A relaxation method for finding a common point of convex sets and its application to the solution of problems in convex programming,” *USSR Computational Mathematics and Mathematical Physics*, no. 7, pp. 200–217, 1967.
- [86] J. Eckstein and D. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Mathematical Programming*, no. 55, 1992.
- [87] E. A. Rakhmanov, E. B. Saff, and Y. M. Zhou, “Minimal discrete energy on the sphere,” *Mathematical Research Letters*, vol. 1, pp. 647–662, 1994.
- [88] E. B. Saff and A. B. J. Kuijlaars, “Distributing many points on a sphere,” *The Mathematical Intelligencer*, vol. 19, no. 1, pp. 5–11, December 1997.
- [89] I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel, “Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle,” *Magnetic Resonance in Medicine*, vol. 64, pp. 554–566, 2010.
- [90] A. Tristan-Vega, C. F. Westin, and S. Aja-Fernandez, “Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging,” *NeuroImage*, vol. 47, pp. 638–650, 2009.